

UFØREDØDELIGHET

av

Trine Marie Ellingsen

MASTEROPPGAVE

for graden

*Master i modellering og dataanalyse
Studieretning finans, forsikring og risiko*

(Master of Science)



*Det matematisk- naturvitenskapelige fakultet
Universitetet i Oslo*

Mai 2009

Forord

Denne oppgaven ble til etter en idé fra ansvarshavende aktuar Magne Nilsen i SpareBank 1 Livsforsikring. Jeg er svært takknemlig for at jeg har fått muligheten til å arbeide med forsikring i praksis ved siden av teori-hverdagen på Blindern. Betydning det har hatt for min utdannelse og utvikling de siste årene er stor. En spesiell takk til Magne Nilsen og Katrine Perger, i tillegg til alle andre flotte kollegaer i aktuaravdelingen, for tre fine år. Det vemodig å takke for seg, men det er allikevel spennende å skulle ta fatt på helt nye utfordringer.

Det er med veiledning fra professor Sven Ove Samuelsen at oppgaven har blitt transformert fra idé til sin nåværende form. Tusen takk til Sven Ove for alle tilbakemeldinger på mine spørsmål, og ikke minst tålmod i sine forklaringer.

Takk til alle Hadelandsvenner som lever opp i hverdagen, til studievenner og lesesalskamerater for skarpsindig humor og en spesiell takk til Hanne for hennes smittende latter og entusiasme.

Den største takk til mine nære og kjære: Simen og mine foreldre.

Trine Marie Ellingsen
Blindern, Mai 2009

Innhold

1	Innledning	1
2	Data	3
2.1	Befolkning	3
2.2	Døde i befolkning	4
2.3	Tilgang til uførhet	5
2.4	Døde uføre	8
2.5	Hvem har rett på uførepensjon?	9
2.6	Tabeller	10
3	Teori	13
3.1	Levetidsanalyse	13
3.1.1	Sannsynlighetsmaksimering	14
3.1.2	Varians og konfidensintervall	15
3.2	Gompertz-fordeling	16
3.2.1	Modellsjekk	18
3.3	Regresjon	19
3.3.1	Poisson-regresjon	19
3.3.2	Modellantagelser	20
3.3.3	Devians	22
3.3.4	Akaike's information criterion	23
4	Resultater Gompertz	25
4.1	Dødelighet i befolkningen	25
4.2	Dødelighet blant uføre	27
5	Resultater regresjon	33
5.1	Effekt av alder ved død	33
5.2	Effekt av tilgangsåar	37
5.3	Effekt av tid fra tilgang	41
5.4	Effekt av periode	45
5.5	Effekt av tilgangsåar og tid fra tilgang	49
5.6	Sammenligning av effekter	54

6	Sensitivitet	55
6.1	Metodikk for befolkningsdataene	55
6.2	Aggregering av data	56
6.3	Sammenligning av metoder	59
6.4	Imputasjoner	61
6.4.1	Faste imputasjoner	61
6.4.2	Uniform trekking	64
6.4.3	Interpolering og bootstrapping	66
6.5	Vurdering av sensitivitet	69
7	Diskusjon	71
7.1	Oppsummering av resultater	71
7.2	Tidligere studier	72
7.3	Konklusjon	73
A	Beregning av forventede observasjonstider	75
B	Konfidensintervaller	77
B.1	Konfidensintervaller for $\hat{\mu}$	77
B.2	Konfidensintervaller for $t\hat{p}_x$	78
B.3	Konfidensintervaller for \tilde{c}	79

Kapittel 1

Innledning

Fokuset er stort på det økte antallet uførepensjonister sett i sammenheng med den demografiske utviklingen som skjer i Norge. For å motvirke effektene av dette, er det en politisk målsetning å redusere tilgangen til uførepensjon og å styrke incentivene til å stå i arbeid. I den forbindelse har endel forskning blitt gjort rundt overgangen fra å være yrkesaktiv og til å bli uførepensjonert. Derimot finnes det et begrenset antall undersøkelser på hvordan dødeligheten til de uføre forholder seg til befolkningens. Spesielt lite er omfanget av årsaksrelaterte undersøkelser av uføredødeligheten.

Dødeligheten blant de uføre kan ses på som et av mange mål på helsetilstanden til de som uførepensjoneres. I arbeidet med strategier for å dempe tilveksten til uførepensjonering er dette et av aspektene som bør vurderes. Det er også trolig at uføredødeligheten kan reflektere konsekvensene av å bli uførepensjonert.

Fra en forsikrer sin side vil også kunnskap om dødeligheten blant de uføre være av interesse. Personer som tegner en forsikring danner nødvendigvis ikke et representativt utvalg av befolkningen. Fenomenet med at enkelte grupper av mennesker har større tendens enn gjennomsnittet til å forsikre seg kalles *seleksjonseffekt*. For at avsetningen til fremtidige utbetalinger skal være korrekt, er det viktig at de estimatene på dødeligheten man benytter i modelleringen, tilsvarende den faktiske dødeligheten i forsikringsbestanden. Det er plausibelt å tro at mye av seleksjonseffekten blant de uføre kan håndteres ved å oppdele uføredødeligheten i ulike årsaker til uførhet. Som forsikrer vet man gjerne hvordan andelene av uføreårsaker fordeler seg i en gitt uførebestand. Med kunnskap om dødeligheten for hver av uføregruppene, kan man dermed oppnå et bedre estimat på den totale dødeligheten for denne bestemte bestanden. Det er en slik antagelse, samt et ønske om å kvantifisere forskjellene, som fungerer som motivasjon for denne oppgaven.

Fra både et samfunns- og forsikringsmessig synspunkt vil det altså være av interesse å se nærmere på dødeligheten blant de uføre. I denne oppgaven vil de uføre inndeles i tre omtrent like store årsaksgrupper. Disse er *psykiske lidelser og adferdsforstyrrelser*, *sykdommer i muskel- og skjelettsystemet og bindevev* og *alle andre uføreårsaker*. Dødeligheten i uføregruppene vil også bli sammenlignet opp mot dødeligheten i den generelle befolkningen. Separate analyser foretas for hvert av kjønnene.

Analysene baseres på datauttrekk fra Arbeids- og Velferdsforvaltningen (NAV) og Statistisk Sentralbyrå (SSB). For å forhindre identifikasjon av individer, er det i uføredataene ikke registrert et eksakt antall dersom antallet døde personer er mellom 1 og 4. For at antallet individer i hver celle fortrinnsvis skal være større enn 4, er datamaterialet utlevert på aggregert nivå. Siden både antallet uførepensjonerte og antallet døde er lavt i yngre aldersgrupper, er også grupperingen ganske grov og av ulik størrelse i forskjellige aldre. Behandling og tilrettelegging av aggregerte data vil derfor også i stor grad omfattes av oppgaven.

Det antas at personene som har tilgang til uførepensjon forblir uføre igjennom hele studien. Det vil altså si at vi ikke tar hensyn til reaktivering blant de uføre, men antar at disse utgjør en beskjeden del av bestanden.

I datauttrekkene mottatt fra NAV inngår kun personer med varig uførepensjon. Dette er noe man må være oppmerksom på når man vurderer analyseresultatene. Det er også verdt å merke seg at uføredataene inngår i befolkningsdataene. Vi sammenligner altså ikke de uføre mot de ikke-uføre, men de uføre mot den totale befolkningen. De uføre utgjør en relativt liten del av befolkningen i ung alder. Ettersom de uføre nærmer seg pensjonsalder begynner de allikevel å utgjøre en betydelig del av befolkningen. På grunn av formen dataene foreligger på, er dette ikke blitt tatt hensyn til i analysene. Behandling av data og alle beregninger er foretatt ved hjelp av statistikkpakken **R**[11].

Av tilgjengelig litteratur rundt emnet, vil masteroppgaven til Paulsboe[10] benyttes som bakgrunnsmateriale for denne oppgaven. En nærmere beskrivelse av hans oppgave kan finnes i kapittel 7.

Oppgaven er organisert som følger:

I kapittel 2 beskriver vi de ulike datasettene som benyttes i oppgaven.

I kapittel 3 gjennomgår vi først litt generell teori om levetidsanalyse. Vi presenterer Gompertz-fordelingen og utleder sannsynlighetsmaksimeringsestimatorene for parameterne i fordelingen. Til regresjonsanalysen trenger vi litt bakgrunnsteori om generaliserte lineære modeller og Poisson-regresjon. Dette gjør vi også rede for i kapittelet. Vi forklarer nærmere om metodikken som benyttes til å måle godheten av tilpasningen til regresjonsmodellen.

I kapittel 4 vises resultatene fra tilpasningen av befolknings- og uføredataene til en Gompertz-fordeling.

I kapittel 5 benyttes Poisson-regresjon til tilpasning av dataene våre. Effekten de ulike forklaringsvariablene har på dødeligheten blir også vurdert.

I kapittel 6 redegjør vi for ulike feilkilder og usikkerhetsmomenter som forekommer i beregningene. Vi forsøker å kvantifisere mulige usikkerhetsmomenter, og vurderer hvilke innvirkninger disse har på resultatene.

I kapittel 7 oppsummerer vi kort resultatene fra analysene og kommenterer på disse. Vi sammenligner også resultatene med andre studier foretatt omkring emnet. Mulige forbedringer og forslag til videre arbeid vil også bli gitt.

Kapittel 2

Data

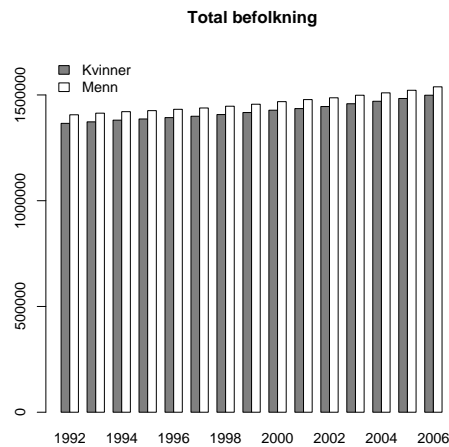
Datamaterialet som benyttes i oppgaven er mottatt fra Arbeids- og Velferdsforvaltningen (NAV) og hentet ut fra Statistisk Sentralbyrå (SSB) sin statistikkbank. Vi skal i inneværende kapittel se nærmere på datagrunnlagene slik at vi kan danne oss et bilde av hvordan befolknings- og uføredataene foreligger. De fire hovedgruppene av data som har blitt benyttet, nemlig befolkningsstørrelse, døde i befolkning, nye uføre og døde uføre, beskrives mer detaljert under.

2.1 Befolkning

De generelle befolkningsdataene, er hentet fra SSB sin statistikkbank og angir befolkningsstørrelsen i Norge per 1. januar i årene fra 1992 til 2006. Et utsnitt av dataene vises i tabell 2.6.1. Folkemengden er splittet i følgende grupperinger, hvor antallet grupper vises i parentes:

Kjønn:	M, K	(2)
Alder:	17, 18, ..., 66	(50)
År:	1992, 1993, ..., 2006	(15)

Eksempelvis har de som er født i 1974 fylt 17 år før 1.1.1992. Disse blir 18 i løpet av kalenderåret 1992. Fra Figur 2.1.1, ser vi at den totale befolkningen i det aktuelle aldersspennet, i perioden fra 1992 til 2006, er jevnt økende for begge kjønn. Vi ser at det er noen flere menn enn kvinner i aldersgruppene vi ser på.



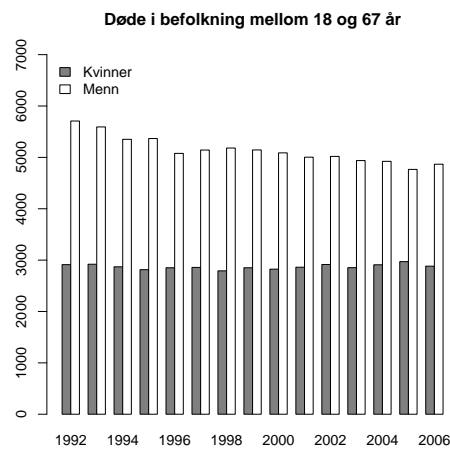
Figur 2.1.1: Total befolkning i alderen 18-67 i årene fra 1992 til 2006, splittet på kjønn.

2.2 Døde i befolkning

Datamaterialet som omfatter de døde i befolkningen, er mottatt fra NAV. Et utsnitt av dataene vises i tabell 2.6.2. Materialet forteller hvor mange innen gitte fødselskohorter som dør i hvert kalenderår fra 1992 til 2006. Dette gir følgende grupperinger:

Kjønn: M, K (2)
Fødselsår: 1925, 1926, ..., 1974 (50)
Dødsår: 1992, 1993, ..., 2006 (15)

Av Figur 2.2.1 ser vi at det totale antallet døde menn er nesten dobbelt så stort som for kvinner i den samme perioden. Antallet døde menn er synkende over perioden, mens antallet døde kvinner ikke kan sies å endres signifikant.



Figur 2.2.1: Totalt antall døde kvinner og menn mellom 18 og 67 år for årene mellom 1992 og 2006.

2.3 Tilgang til uførhet

Datamaterialet for tilgang til uførhet og antallet døde blant de uførepensjonerte, er også mottatt fra NAV. Et utsnitt av disse dataene vises i tabell 2.6.3. Som tidligere nevnt er uføredataene inndelt i tre forskjellige årsaksgrupper. Lidelsene som inngår i årsaksgruppene er listet under. Notasjonen som benyttes i oppgaven vises i parentes:

Psykiske lidelser og adferdsforstyrrelser (Psykiske lidelser, **P**):

- Organiske psykiske lidelser/schizofrene og paranoide lidelser
- Affektive lidelser (stemningslidelser)
- Nevroser og adferds- og personlighetsforstyrrelser
- Psykiske lidelser på grunn av rusmiddel- eller tablettbruk
- Psykisk utviklingshemming

Sykdommer i muskel-skjelettsystemet og bindevev (Muskelsykdommer, **M**):

- Reumatoid artritt
- Artroser
- Ryggsykdommer
- Uspesifisert reumatisme/myalgi

De andre årsakene (**A**) domineres av følgende lidelser:

- Sykdommer i sirkulasjonssystemet
- Sykdommer i nervesystemet
- Svulster
- Skader, forgiftninger, og vold

Grupperingen er som følger:

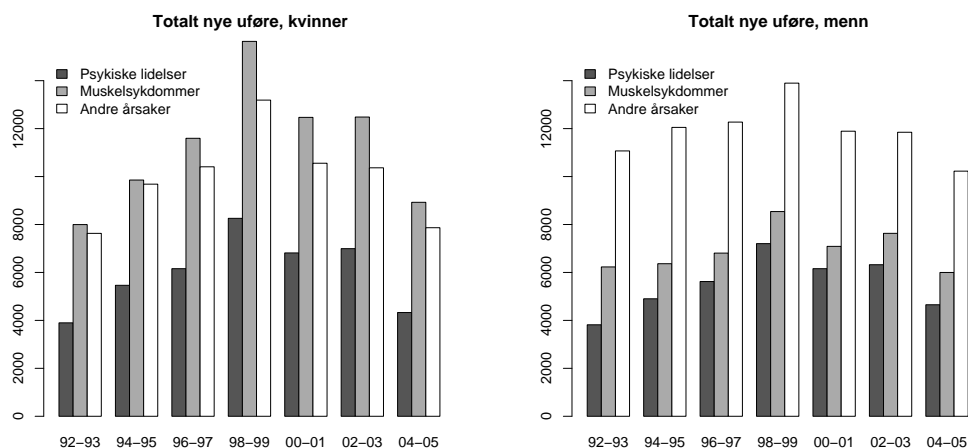
Kjønn: M, K (2)

Alder: 18-25, 26-30, 31-35, 36-40, 41-45, 46-48,
49-51, 52-53, 54-55, 56-57, 58-59, 60-61, (15)
62-63, 64-65, 66-67

Tilgangså: 1992,...,2006 (15)

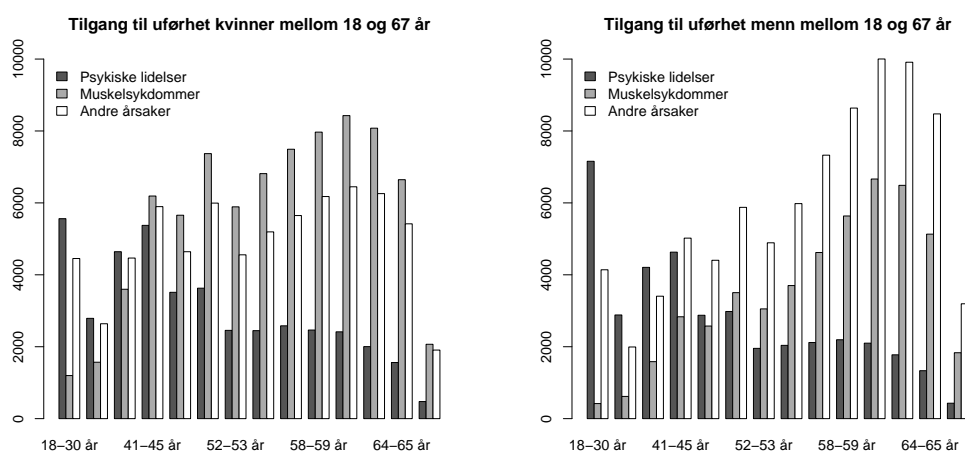
Uføreårsak: P, M, A (3)

Av tilgangen til uførhet, ser vi fra figur 2.3.1 en økning i antallet nye uføre fram til begynnelsen av dette århundret, da tilgangen begynte å avta. Det virker imidlertid som om forholdet mellom antallet nye uføre innen hver gruppe holder seg relativt stabilt over perioden. Alle uføregruppene ser ut til å ha den samme trenden. Mens muskelsykdommer er årsaksgruppa som utgjør hovedvekten av nye uførepensjonerte kvinner, domineres uførepensjoneringen blant menn av gruppa av andre årsaker.



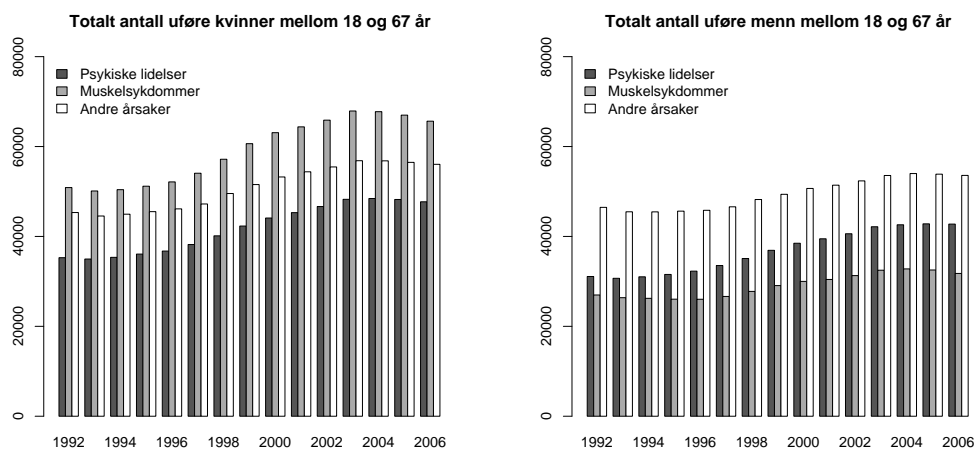
Figur 2.3.1: Antallet nye uføre i årene fra 1992-1993 til 2004-2005 for henholdsvis kvinner og menn.

I figur 2.3.2 vises den totale tilgangen til uførhet i årene fra 1992 til og med 2005 fordelt på alder ved tilgang. Vi ser at psykiske lidelser er den uføreårsaken som bidrar med den største tilgangen i ung alder. Tilgangen til uførhet grunnet denne lidelsen er også avtagende ettersom alderen øker. Det motsatte er tilfellet for gruppa uføretrygdet grunnet muskelsykdommer, hvor antallet som uføretrygdes av denne årsaken er økende med alder. For kvinnene er muskelsykdommer den dominerende uføreårsaken fra de er i alder 40 år og eldre. For menn domineres derimot tilgangen i disse aldre av andre årsaker.



Figur 2.3.2: Totalt antall nye uføre i årene fra 1992-1993 til 2004-2005 fordelt på aldersgrupper henholdsvis kvinner og menn.

Også det totale antallet uføre i hver av uføregruppene er mottatt fra NAV, men benyttes ikke videre til analyser. For å kunne danne seg et bilde av hvordan utviklingen er i tidsperioden vi ser på, kan det allikevel være nyttig å se nærmere på disse dataene. Det vises av figur 2.3.3 at det er flere uførepensjonerte kvinner enn menn. Vi ser at den største årsaksgruppen hos kvinner er muskelsykdommer, mens det blant menn er andre årsaker. Dette forholdet så vi også av tilgangen til uførhet. Det er flere menn uførepensjonert grunnet psykiske lidelser, enn det er av muskelsykdommer. Derimot har menn en større tilgang til uførepensjon i gruppa av muskelsyke enn i gruppa med psykiske lidelser. Foruten dette, er forholdet mellom årsakene ved tilgang og i den totale uføregruppa likt. Det er også tydelig at det forekommer en økning i det totale antallet uførepensjonister i årene fram til 2004. Veksten stagnerer da blant menn, og for kvinner får vi en negativ vekst i det totale antallet uførepensjonister.



Figur 2.3.3: Totalt antall uføre i årene fra 1992 til 2006 for henholdsvis kvinner og menn.

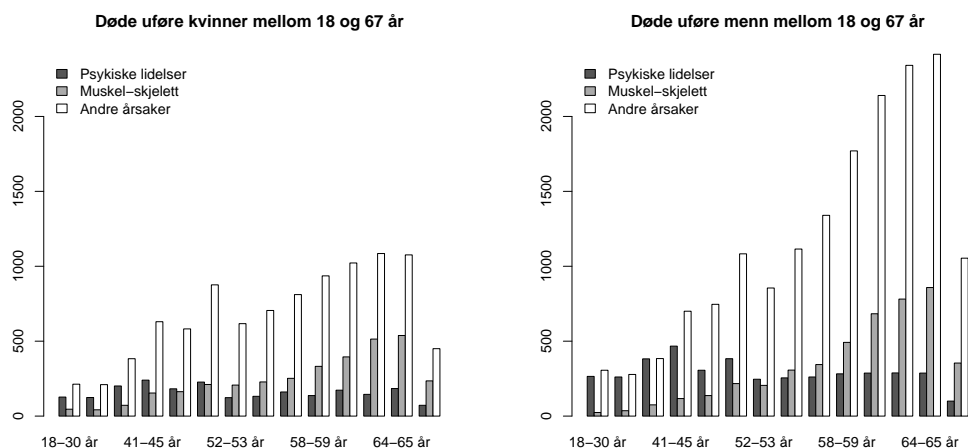
2.4 Døde uføre

Et utsnitt av dataene med antallet døde uføre vises i tabell 2.6.4. På grunn av at små datamengder er tilgjengelig, gis antallet døde uføre ut fra parvise tilgangsårs og død i perioder på tre og tre år. Dersom antallet døde uføre er mellom 1 og 4, er det ikke registrert et eksakt antall. I stedet indikeres disse med en stjerne(*) i datafilen. Dette er i hensyn til personvernet. Grupperingene blir som følger:

Kjønn:	M, K	(2)
Alder:	18-30, 31-35, 36-40, 41-45, 46-48, 49-51, 52-53, 54-55, 56-57, 58-59, 60-61, 62-63, 64-65, 66-67	(14)
Tilgangsårs:	1992-1993, 1994-1995, ..., 2004-2005	(7)
Dødsår:	1992-1994, 1995-1997, ..., 2004-2006	(5)
Uføreårsak:	P, M, A	(3)

Overlappingen mellom tilgangsårene og dødsårene er av ulik lengde, så derfor gir det liten mening å framstille dette grafisk. Derimot kan vi få et inntrykk av hvordan antallet døde uføre fordeler seg mellom de ulike aldersgruppene ved å summere antall døde fra samme aldersgruppe ved tilgang, men med ulike tilgangs- og dødsår. En slik framstilling vises i figur 2.4.1. Antallet døde blant de uførepensjonerte grunnet psykiske lidelser holder seg stabilt med økning i alder, men er en anelse avtagende for menn. Derimot ser vi en økning

i antallet dødsfall blant de muskelsyke for begge kjønn og en kraftig vekst i antallet døde uførepensjonerte grunnet andre årsaker, spesielt for menn.



Figur 2.4.1: Antallet døde uføre i de ulike aldersgruppene for kvinner og menn. * er her satt til å være 3 døde.

2.5 Hvem har rett på uførepensjon?

I følge Lov om folketrygd[2], se kap. 12, er formålet med uførepensjon “å sikre inntekt for personer som har fått sin inntektsevne eller arbeidsevne varig nedsatt på grunn av sykdom, skade eller lyte”. For å ha rett til å motta uføreytelser, sier Lov om Folketrygd at følgende fem kriterier må være tilfredsstilt:

- Medlemskap i folketrygden de siste tre årene fram til uføretidspunktet.
- Alder mellom 18 og 67 år.
- Inntektsevnen til vedkomne må være nedsatt på grunn av varig sykdom, skade eller lyte. Det er også et krav at sykdommen, skaden eller lytet er hovedårsaken til den nedsatte inntektsevnen.
- Vedkommende må ha gjennomgått hensiktsmessig behandling for å bedre arbeidsevnen og ha forsøkt å gjennomføre et individuelt og hensiktsmessig attføringsopplegg.
- Evnen til å utføre inntektsgivende arbeid er varig nedsatt med minst halyparten.

Fra 1.1.2004 ble uføreytelser fra folketrygden splittet i varig uførepensjon og tidsbegrenset uførestønad. Den tidsbegrensede uførestønaden utbetales i følge Lov om Folketrygd “dersom det etter en helhetsvurdering er sannsynlig at inntektsevnen kan bedres”. Retten til en slik stønad gis for en periode på inntil fire år. Personene som innvilges tidsbegrenset uførestønad må allikevel oppfylle de ovennevnte kriteriene for å kunne motta uføreytelser.

2.6 Tabeller

Kvinner		År		
Alder	1992	1993	1994	
17 år	30018	28155	26840	
18 år	30436	30134	28372	
19 år	31923	30478	30223	
20 år	32531	32084	30660	
21 år	32047	32666	32227	
22 år	33539	32184	32888	
23 år	33409	33750	32411	

Tabell 2.6.1: *Utsnitt fra datasettet for befolkningsstørrelse. Kvinner.*

Kvinner		Fødselsår				
Dødsår	1925	1926	1927	1928	1929	
1992	239	226	187	191	149	
1993	254	261	242	188	179	
1994	271	231	216	213	178	
1995	322	297	236	209	202	
1996	331	321	247	231	195	

Tabell 2.6.2: *Utsnitt fra datasettet for antallet døde i befolkningen. Kvinner*

Kvinner		Alder 18-25 år		
Tilgangår	Psykiske lidelser	Muskelsykdommer	Andre årsaker	
1992	81	6	92	
1993	111	5	104	
1994	169	21	133	
1995	182	23	148	
1996	162	22	141	

Tabell 2.6.3: *Utsnitt fra datasettet for tilgang til uførepensjon. Kvinner.*

Kvinner		Alder 18-30 år		
Tilgangår	Dødsår	Psykiske lidelser	Muskelsykdommer	Andre årsaker
92-93	92-94	*		15
	95-97	5		*
	98-00	*		*
	01-03	6	*	8
	04-06	6	*	7

Tabell 2.6.4: *Utsnitt fra datasettet for antallet døde blant de uføre. Kvinner. Blanke felter indikerer ingen døde, mens * angir mellom 1 og 4 døde.*

Kapittel 3

Teori

3.1 Levetidsanalyse

Ifra dataene observerer vi tiden fra en person går inn i studien og fram til personen forlater studien. Disse observasjonstidene vil for befolkningsdataene være tiden fra en person er i en alder x og til personen dør. For de uføre vil alderen x , som vi starter å observere personen ved, tilsvare alderen denne personen har ved uførepensjonering. Siden observasjonen foregår over et begrenset tidsperspektiv, vil vi ikke kunne observere fullstendige levetider for alle individene som inngår i studien. Vi sier derfor at dataene er *høyresensurerte* og kaller tiden da studien avsluttes for *sensurtiden*. Siden individene har ulike alder da studien starter, vil dataene også være det vi kaller *venstretrunkerte*. Observert levetid utover alder x benevnes T_x , og tilsvare tiden fra x og fram til hva som inntreffer først av dødsfall og sensurtid. For å holde rede på hvilke levetider som er sensurerte og hvilke som angir faktiske dødstidspunkter, innføres det for hvert individ i en sensurindikator, δ_i . Der

$$\delta_i = \begin{cases} 1 & \text{hvis individ } i \text{ dør underveis i studien,} \\ 0 & \text{hvis individ } i \text{ fortsatt er i live når studien avsluttes.} \end{cases}$$

Definisjon 3.1.1 *La den tilfeldige framtidige levetidsvariabelen T_x ha en kumulativ fordelingsfunksjon $F_x(t) < 1$ for alle t , med tilhørende tetthetsfunksjon $f_x(t)$. Dødsintensiteten $\mu(x+t)$ defineres da som*

$$\mu(x+t) = \frac{f_x(t)}{1 - F_x(t)}. \quad (3.1)$$

Dødsintensiteten $\mu(x+t)$ kan ses på som den umiddelbare dødsrisikoen, gitt overlevelse fram til alder $x+t$. Fra (3.1) ser vi at $\mu(x+t) = -\frac{d}{dt} \ln[1 - F_x(t)]$, og fordi $F_x(0) = 0$ vil

$$F_x(t) = 1 - \exp\left(-\int_0^t \mu(x+\tau) d\tau\right)$$

og

$$f_x(t) = \mu(x+t) \exp\left(-\int_0^t \mu(x+\tau) d\tau\right). \quad (3.2)$$

For fordelingsfunksjonen $F_x(t)$, skal vi videre i oppgaven benytte den aktuarielle betegnelsen ${}_tq_x$, som gir sannsynligheten for at en person ved alder x dør innen t år. Tilsvarende benevner vi overlevelsessannsynligheten som

$$\begin{aligned} {}_tp_x &= 1 - F_x(t) \\ &= \exp\left(-\int_0^t \mu(x + \tau) d\tau\right). \end{aligned} \quad (3.3)$$

3.1.1 Sannsynlighetssmaksimering

Anta nå at fordelingsfunksjonen, og dermed også dødsintensiteten, bestemmes av et sett av parametre θ . Likelihoodbidraget fra et individ i med observert dødstidspunktet t_i , vil være $f_x(t_i; \theta)$. For de sensurerte individene, vil likelihoodbidraget være sannsynligheten for å overleve sensurtiden, nemlig $1 - F_x(t_i; \theta)$. Den totale likelihooden $L(\theta)$ kan ved hjelp av (3.2) og (3.3) skrives som

$$L(\theta) = \prod_{i=1}^n \mu(x_i + t_i; \theta)^{\delta_i} \exp\left(-\int_0^{t_i} \mu(x_i + \tau; \theta) d\tau\right). \quad (3.4)$$

Definisjon 3.1.2 *La oss anta at $t_0 = 0 < t_1 < \dots < t_k = K$ er partisjoner av studiens tidsintervall $[0, K]$ og at dødsintensiteten er stykkevis konstant, og spesielt konstant i hvert av de K delintervallene. Vi definerer dødsintensiteten som*

$$\mu_j = \mu(x_i + t_i; \theta),$$

dersom $x_i + t_i$ inngår i delintervall j .

Vi benytter også notasjonen $\mu(x_i + t_i - k; \theta) = \mu_{j-k}$ for alle $k \leq t_i$. Videre lar vi $\mu = (\mu_1, \mu_2, \dots, \mu_k)$ være vektoren av slike dødsintensiteter. Antallet hendelser for individ i i delintervall j benevnes D_{ij} , og total tid individ i er til risiko i delintervall j benevnes T_{ij} .

Likelihooden for individ i kan nå skrives som et produkt over alle likelihoodbidragene fra hvert av delintervallene j ,

$$L_i(\mu) = \prod_{j=1}^K \mu_j^{D_{ij}} \exp(-T_{ij}\mu_j).$$

Den totale likelihooden blir da

$$\begin{aligned} L(\mu) &= \prod_{i=1}^n \prod_{j=1}^K \mu_j^{D_{ij}} \exp(-T_{ij}\mu_j) \\ &= \prod_{j=1}^K \mu_j^{D_j} \exp(-T_j\mu_j) \\ &= \prod_{j=1}^K L_{(j)}(\mu_j), \end{aligned} \quad (3.5)$$

hvor $L_{(j)}(\mu_j)$ er likelihoodbidraget fra intervall j og vi har at

$$D_j = \sum_i D_{ij} = \text{Totalt antall døde i intervall } j$$

$$T_j = \sum_i T_{ij} = \text{Totalt antall personår til risiko i intervall } j.$$

Log-likelihooden blir

$$l(\boldsymbol{\mu}) = \log L(\boldsymbol{\mu})$$

$$= \sum_{j=1}^K D_j \log(\mu_j) - T_j \mu_j$$

med

$$\frac{\partial}{\partial \mu_j} l(\boldsymbol{\mu}) = \frac{D_j}{\mu_j} - T_j, \quad \text{og} \quad (3.6)$$

$$\frac{\partial^2}{\partial \mu_i \partial \mu_j} l(\boldsymbol{\mu}) = \begin{cases} -\frac{D_j}{\mu_j^2}, & i = j \\ 0, & i \neq j. \end{cases} \quad (3.7)$$

Fra dette får vi de følgende to resultater:

Proposisjon 3.1.3 *Sannsynlighetsmaksimeringsestimatoren (SME) for μ_j i intervall j er gitt ved*

$$\hat{\mu}_j = \frac{D_j}{T_j}, \quad (3.8)$$

og benevnes årlig rate.

Proposisjon 3.1.4 *Overlevelsessannsynligheten innsatt SME er gitt ved*

$${}_t\hat{p}_x = \exp \left(- \sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}} \right). \quad (3.9)$$

3.1.2 Varians og konfidensintervall

Definisjon 3.1.5 *Matrisen $\bar{\mathbf{J}}(\boldsymbol{\mu})$, bestående av elementer*

$$\bar{\mathbf{J}}(\boldsymbol{\mu})_{ij} = -\frac{\partial^2}{\partial \mu_i \partial \mu_j} l(\boldsymbol{\mu}),$$

kalles den observerte informasjonsmatrisen, og $\bar{\mathbf{J}}^{-1}(\boldsymbol{\mu})$ kalles kovariansmatrisen.

Kovariansmatrisen $\bar{\mathbf{J}}^{-1}(\boldsymbol{\mu})$ innsatt SME $\hat{\mu}_j$ fra (3.8), blir bestående av elementer

$$\bar{\mathbf{J}}^{-1}(\hat{\boldsymbol{\mu}})_{ij} = \begin{cases} \frac{D_j}{T_j^2}, & i = j \\ 0, & i \neq j. \end{cases}$$

Ethvert element μ_j av $\boldsymbol{\mu}$ vil, for et tilstrekkelig stort antall observasjoner, ha en fordelingsfunksjon som er tilnærmet normalfordelt. Dersom vi setter inn $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$, vil vi altså ha at

$$\hat{\mu}_j \sim N(\mu_j, \hat{\sigma}_{jj}^2),$$

hvor $\hat{\sigma}_{jj}^2 = \{\bar{\mathbf{J}}^{-1}(\hat{\boldsymbol{\mu}})\}_{jj}$ er det j 'te elementet i diagonalen til kovariansmatrisen.

Siden tilgangen på data er begrenset, vil ikke alltid antagelsen om at $\hat{\mu}_j$ er normalfordelt være korrekt. Tilpasningen av konfidensintervallene blir bedre dersom vi i stedet antar at $\log \hat{\mu}_j$ er normalfordelt.

Proposisjon 3.1.6 *Et tilnærmet $(1 - \alpha)100\%$ konfidensintervall for $\hat{\mu}_j$ vil være*

$$\frac{D_j}{T_j} \exp\left(\pm z(\alpha/2) \sqrt{\frac{1}{D_j}}\right), \quad (3.10)$$

hvor $z(\alpha/2)$ er α -kvantilen i en standard normalfordeling.

Proposisjon 3.1.7 *Et tilnærmet $(1 - \alpha)100\%$ konfidensintervall for ${}_t\hat{p}_x$ vil være*

$$\exp\left(-\sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}}\right) \pm z(\alpha/2) \exp\left(-\sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}}\right) \sqrt{\sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}^2}}. \quad (3.11)$$

Utleddningen av konfidensintervallene vises i tillegg B.1 og B.2.

3.2 Gompertz-fordeling

Gompertz-Makehams fordeling er en mye benyttet fordeling ved dødlighetsmodellering innen forsikring. I denne oppgaven benyttes en noe forenklet fordeling, nemlig en ren Gompertz-fordeling, som en av metodene for å analysere dødelighet. Se Gerber[4] kap. 2.3.

Definisjon 3.2.1 *Dødsintensiteten ved alder $x + t$ er ved Gompertz-Makeham-fordelingen definert som*

$$\mu_{x+t} = \alpha + \beta c^{x+t}.$$

En ren Gompertz-intensitet oppnås ved å sette $\alpha = 0$,

$$\mu_{x+t} = \beta c^{x+t}. \quad (3.12)$$

Overlevelsessannsynligheten fra (3.3) blir under Gompertz-fordelingen

$$\begin{aligned} {}_t p_x &= \exp \left(- \int_0^t \mu_{x+\tau} d\tau \right) \\ &= \exp \left(-\beta \frac{c^x (c^t - 1)}{\log(c)} \right). \end{aligned} \quad (3.13)$$

Den kumulative fordelingsfunksjonen blir da

$$\begin{aligned} {}_t q_x &= 1 - {}_t p_x \\ &= 1 - \exp \left(-\beta \frac{c^x (c^t - 1)}{\log(c)} \right), \end{aligned}$$

med tetthetsfunksjonen

$$\frac{d}{dt} {}_t q_x = \beta c^{x+t} \exp \left(-\beta \frac{c^x (c^t - 1)}{\log(c)} \right).$$

For å estimere parameterne β og c for et gitt datasett, benyttes likelihooden gitt ved (3.4). Likelihoodbidragene fra de sensurerte individene vil også her tilsvare sannsynligheten for overlevelse fram til sensurtidspunktet. For de individene hvor et reelt dødstidspunkt observeres, vil bidraget være sannsynlighetstettheten på dødstidspunktet. Dette gir likelihooden

$$L(\beta, c) = \prod_{i=1}^n (\beta c^{x_i+t_i})^{\delta_i} \exp \left(-\beta \frac{c^{x_i} (c^{t_i} - 1)}{\log(c)} \right). \quad (3.14)$$

Log-likelihooden blir da

$$\begin{aligned} l(\beta, c) &= \log(\beta) \sum_{i=1}^n \delta_i \\ &\quad + \log(c) \sum_{i=1}^n \delta_i (t_i + x_i) - \frac{\beta}{\log(c)} \sum_{i=1}^n c^{x_i} (c^{t_i} - 1). \end{aligned}$$

Ved å sette den partiellderiverte med hensyn på β lik null og løse, får vi et estimeringsresultat $\hat{\beta}$ for β , gitt c :

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \frac{1}{\beta} \sum_{i=1}^n \delta_i - \frac{1}{\log(c)} \sum_{i=1}^n c^{x_i} (c^{t_i} - 1) = 0 \\ \hat{\beta}(c) &= \frac{\log(c) \sum_{i=1}^n \delta_i}{\sum_{i=1}^n c^{x_i} (c^{t_i} - 1)}. \end{aligned}$$

Den partiellderiverte med hensyn på c , hvor $\hat{\beta}$ innsettes for β blir dermed

$$\begin{aligned} f'(c) &= \frac{\partial l(\hat{\beta}(c), c)}{\partial c} = \frac{1}{c} \sum_{i=1}^n \delta_i (t_i + x_i) - \frac{\beta}{c \log(c)} \sum_{i=1}^n x_i c^{x_i} (c^{t_i} - 1) \\ &\quad + \frac{\beta}{c \log^2(c)} \sum_{i=1}^n c^{x_i} (c^{t_i} - 1) - \frac{\beta}{\log(c)} \sum_{i=1}^n c^{x_i} t_i c^{t_i-1} \\ &= \frac{1}{c} \sum_{i=1}^n \delta_i (t_i + x_i) - \frac{\sum_{t=1}^n \delta_i \sum_{i=1}^n c^{x_i} t_i c^{t_i-1}}{\sum_{i=1}^n c^{x_i} (c^{t_i} - 1)} \\ &\quad + \frac{1}{c \log(c)} \sum_{i=1}^n \delta_i - \frac{\sum_{t=1}^n \delta_i \sum_{i=1}^n x_i c^{x_i-1} c^{t_i} - 1}{\sum_{i=1}^n c^{x_i} (c^{t_i} - 1)}. \end{aligned}$$

Et analytisk uttrykk kan ikke bli funnet for \hat{c} , så en løsning av likningen $f'(c) = 0$ må finnes numerisk. Dette kan enten gjøres direkte ved hjelp innebygde funksjoner for løsning av likninger i \mathbf{R} som *uniroot*, eller ved å benytte Newton-Raphsons metode.

Proposisjon 3.2.2 *Newton-Raphsons metode er en iterativ prosedyre, som finner den c som gir $f'(c) = 0$ som følger:*

$$c^{(s+1)} = c^{(s)} - \frac{f'(c^{(s)})}{f''(c^{(s)})}. \quad (3.15)$$

Prosedyren starter med initielt gjett på c , og gjentas til endringen fra $c^{(s)}$ til $c^{(s+1)}$ er mindre enn en forhåndsbestemt liten størrelse.

3.2.1 Modellsjekk

For å sjekke antagelsene om at dataene kan tilpasses en Gompertz-fordeling, evalueres parametertilpasningen til c . Parameteren c uttrykker den konstante økningen i dødsintensiteten i Gompertz-fordelingen,

$$c = \frac{\mu_{j+1}}{\mu_j}.$$

Veksten fra år j til år $j + 1$ uttrykkes ved de årlige ratene som \tilde{c}_j ,

$$\tilde{c}_j = \frac{\hat{\mu}_{j+1}}{\hat{\mu}_j},$$

hvor $\hat{\mu}_j$ er SME for μ_j fra (3.8).

Proposisjon 3.2.3 *Et tilnærmet $(1 - \alpha)100\%$ konfidensintervall for \hat{c} er gitt ved*

$$\left(\frac{T_j D_{j+1}}{D_j T_{j+1}} \right) \exp \left(\pm z(\alpha/2) \left(\sqrt{\frac{1}{D_{j+1}} + \frac{1}{D_j}} \right) \right). \quad (3.16)$$

Utleddningen av konfidensintervallet finnes i tillegg B.3.

Som en modellsjekk kan vi verifisere at den optimale tilpasningen \hat{c} , funnet fra (3.15), er inneholdt i dette konfidensintervallet for et passende valg av α .

3.3 Regresjon

Dødeligheten til en person kan forklares av en sammensetning av en rekke ulike faktorer som beskriver denne personen. For enkelt å kunne se effekten av de ulike faktorene, vil det være naturlig å foreta regresjonsanalyser på dataene. Antallet døde i et gitt tidsintervall blir da vår responsvariabel. Antallet personår til risiko inngår i regresjonen som en kjent størrelse, og kalles et *offset*. Responsvariabelen blir beskrevet av faktorene som innvirker på levetiden, disse kalles forklaringsvariabler.

Med responsvariabel Y_i får vi en klassisk lineær modell med relasjonen

$$Y_i = z_i' \beta + \epsilon_i, \quad i = 1, \dots, n,$$

hvor z_i kalles designvektor, en funksjon av forklaringsvariablene x_i . Enklest uttrykkes z_i på formen $z_i' = z(x_i) = (1, x_i')$. Feilleddene ϵ_i antas å være normalfordelte og uavhengige

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n.$$

Vi kan derfor si at responsvariabelen har fordelingen

$$Y_i \sim N(\lambda_i, \sigma^2), \quad i = 1, \dots, n.$$

For å bevare tidligere notasjon, benevnes responsens forventning som

$$\lambda_i = E[Y_i] = z_i' \beta, \quad i = 1, \dots, n.$$

Det er ikke alltid opplagt at responsvariabelen er normalfordelt, så en videre klasse av modeller benyttes. Disse kalles generaliserte lineære modeller (GLM). Se McCullagh and Nelder[9] kap. 2.2.

Proposisjon 3.3.1 *GLMer karakteriseres ved følgende tre egenskaper:*

1. Y_1, \dots, Y_n uavhengige responsvariabler som alle kommer fra samme eksponentielle klasse av fordelinger.
2. Den lineære prediktoren $\eta = z' \beta$ er en lineær sum av av effektene til en eller flere av forklaringsvariablene.
3. Link-funksjonen $g(\lambda) = \lambda = z' \beta$ som relaterer forventningen til responsen Y .

3.3.1 Poisson-regresjon

Til tilpasningen av dataene i oppgaven benyttes en generalisert lineær modell med log-lineære, Poisson-fordelte responser. Det vil altså si at betingelsene for en GLM oppfylles som følger:

1. Y_1, \dots, Y_n uavhengige responsvariabler, $Y_i \sim \text{Pois}(\lambda_i)$ for $i = 1, \dots, n$.
2. Lineær prediktor $\eta = z' \beta$.

3. Link-funksjonen $g(\lambda) = \log(\lambda) = \eta$.

La oss benevne en likelihood $L^*(\boldsymbol{\mu})$, hvor vi antar Poisson-fordelte responser med forventning $\lambda_j = T_j \mu_j$ og stykkevis konstante rater,

$$L^*(\boldsymbol{\mu}) = \prod_{j=1}^K \left\{ \frac{(T_j \mu_j)^{D_j}}{D_j!} \exp(-T_j \mu_j) \right\}. \quad (3.17)$$

I sammenligning med likelihooden til høyresensurerte data under eksponentialfordelingen, $L(\boldsymbol{\mu})$ fra (3.5), får vi følgende resultat:

Proposisjon 3.3.2 *Likelihooden under eksponentialfordelingen, $L(\boldsymbol{\mu})$, er proporsjonal med Poisson-likelihooden $L^*(\boldsymbol{\mu})$,*

$$\begin{aligned} L(\boldsymbol{\mu}) &= L^*(\boldsymbol{\mu}) \prod_{j=1}^K \frac{D_j!}{T_j^{D_j}} \\ &\propto L^*(\boldsymbol{\mu}), \end{aligned}$$

siden D_j og T_j ikke avhenger av $\boldsymbol{\mu}$.

Det er derfor mulig å benytte Poisson-regresjon til å optimere likelihooden, uten å måtte foreta antagelser om at responsen er Poisson-fordelt. Se Aalen et al.[1] kap. 5.2.1.

3.3.2 Modellantagelser

Regresjonen foretas på ulike sammensetninger av forklaringsvariabler. I oppgaven benyttes følgende forklaringsvariabler, med tilhørende notasjon

Kovariat	Forklaring
ÅRSAK	Årsak til uførhet
KJØNN	Kjønn
ADØD	Alder ved død
TILGANG	Tilgangssår
PERIODE	Død i periode
TID FRA TILG	Tid fra tilgang

Separate analyser foretas på forklaringsvariablene kjønn og uføreårsak. En kategorisk forklaringsvariabel i med k mulige kategorier, $1, \dots, k$ gis ved en vektor \mathbf{z}_i med $q=k-1$ komponenter. $\mathbf{z}_i = (z_i^{(1)}, \dots, z_i^{(q)})$, der

$$z_i^{(j)} = \begin{cases} 1 & \text{hvis kategori } j \text{ er observert, } j=1, \dots, q \\ 0 & \text{ellers.} \end{cases}$$

Effektene fra forklaringsvariablene i modellen virker multiplikativt inn på hverandre.

Eksempel 3.3.3 Dødsintensiteten til personer med alder a og tilgang i år t , skrives som

$$\begin{aligned}\mu_{a,t} &= \exp(\beta_0) \exp(\beta_a) \exp(\beta_t) \\ &= \exp(\beta_0 + \beta_a + \beta_t).\end{aligned}$$

I denne modellen over benyttes forklaringsvariablene alder ved død og tilgangså i regresjonen. Modellbetegnelsen $\text{ADØD} + \text{TILGANG}$ benyttes for henviser til en slik modell.

Kategorien k kaller vi referansekategorien. Dersom denne observeres er \mathbf{z}_i nullvektoren.

Eksempel 3.3.4 Den forventede dødsintensiteten til en person som faller i denne referansekategorien skrives som

$$\mu_{k,k} = \exp(\beta_0).$$

Eksempel 3.3.5 De årlige ratene funnet i (3.8) blir et spesialtilfelle av Poisson-regresjonen med modellen ADØD . Dødsintensiteten forklares da kun av alder ved død $\mu_a = \exp(\beta_0 + \beta_a)$, med et tilhørende konfidensintervall gitt ved (3.10).

Vi har også modeller hvor vi ønsker å se effekten av interaksjoner mellom forklaringsvariablene, og uttrykker dette med $*$ i modellbetegnelsen.

Eksempel 3.3.6 Den samme modellen, men hvor vi nå også angir et ledd for interaksjonen vil være gitt med dødsintensiteten

$$\mu_{a,t,at} = \exp(\beta_0 + \beta_a + \beta_t + \beta_{a,t}),$$

der $\beta_{a,t}$ angir interaksjonseffekten mellom alder ved død a og tilgang i år t . Dette gir modellen $\text{ADØD} + \text{TILGANG} + \text{TILGANG} * \text{ADØD}$. Den noe mer komprimerte notasjonen som benyttes for dette er $\text{TILGANG} * \text{ADØD}$.

Proposisjon 3.3.7 Den relative risikoen (RR) er gitt som forholdstallet mellom dødsintensiteten i ulike grupper mot en bestemt referansegruppe.

Eksempel 3.3.8 Den relative risikoen for død i tilgangså t relativt til død i samme år som uførepensjonering blir da

$$\widehat{RR}_t = \frac{\hat{\mu}_{a,t}}{\hat{\mu}_{a,0}} = \exp(\beta_t),$$

der t er antall år fra uførepensjonering. Konfidensintervallet for \widehat{RR}_t blir

$$\exp(\beta_t \pm z(\alpha/2)\sigma_{\beta_t}),$$

der σ_{β_t} er standardavviket til koeffisienten β_t , og $z(\alpha/2)$ er $\alpha/2$ -kvantilen i en standard normalfordeling.

3.3.3 Devians

Det er ønskelig å måle avvikene fra den tilpassede modellen og til dataene, for å se hvor unøyaktig modellen er. For generaliserte lineære modeller er devians et mye benyttet mål på dette avviket. Se McCullagh and Nelder[9] kap. 2.3.

Definisjon 3.3.9 *Deviansen, \mathcal{D} , mellom den mettede modellen og modellen vi vurderer, er gitt ved*

$$\mathcal{D}(\hat{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}) = -2[l^*(\hat{\boldsymbol{\mu}}) - l^*(\tilde{\boldsymbol{\mu}})],$$

hvor $l^*(\tilde{\boldsymbol{\mu}})$ er log-likelihooden for den mettede modellen og $l^*(\hat{\boldsymbol{\mu}})$ er log-likelihooden for den aktuelle modellen.

Den mettede modellen er en modell som er perfekt tilpasset dataene, altså har en parameter for ethvert datapunkt. Siden vi har aggregerte data, vil mange av observasjonene har sammenfallende datapunkter. Vi benevner en vektor $\tilde{\boldsymbol{\mu}}$ av lengde K , bestående av slike tilpassede dødsintensiteter for hvert av de sammenfallende datapunktene. Denne modellen vil være perfekt tilpasset, men svært lite informativ da den ikke gir noen sammenfatning av dataene. Derimot kan en slik modell benyttes til å måle avvik til modeller med færre parametre. For en forenklet modell, hvor kun noen av forklaringsvariablene vurderes, benevnes en vektor av dødsintensiteter $\hat{\boldsymbol{\mu}}$ av lengde K . For et hvert element $\hat{\mu}_j$ i $\hat{\boldsymbol{\mu}}$ vil alle individer da ha sammenfallende kombinasjon j , av forklaringsvariabler. Fra (3.17) har vi likelihood hvor vi antar Poisson-fordelte responser, så log-likelihooden til den mettede modellen blir da

$$l^*(\tilde{\boldsymbol{\mu}}) = \sum_{j=1}^K D_j \log(T_j \tilde{\mu}_j) - T_j \tilde{\mu}_j - \log(D_j!). \quad (3.18)$$

Log-likelihooden for den forenklede modellen blir

$$l^*(\hat{\boldsymbol{\mu}}) = \sum_{j=1}^K D_j \log(T_j \hat{\mu}_j) - T_j \hat{\mu}_j - \log(D_j!). \quad (3.19)$$

Proposisjon 3.3.10 *For Poisson-fordelte responser blir dermed deviansen*

$$\mathcal{D}(\hat{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}) = 2 \sum_{j=1}^K D_j \log \left(\frac{\tilde{\mu}_j}{\hat{\mu}_j} \right) - T_j (\tilde{\mu}_j - \hat{\mu}_j). \quad (3.20)$$

La oss anta vi har to nestede modeller, hvor vi spesielt har at modell 2 inneholder settet av parametre i modell 1 i tillegg til ytterligere k parametre. Under antagelser om at modell 1 er sann, vil da differansen i deviansen mellom de to modellene være χ^2 -fordelt med k frihetsgrader. Dette resultatet skal vi benytte til å se hvorvidt vi får signifikante bedringer i tilpasningen av modellen ved å tilpasse ytterligere forklaringsvariabler.

3.3.4 Akaike's information criterion

For å kunne sammenligne modeller med et ulikt antall frihetsgrader, benyttes Akaike's information criterion. Se Fahrmeir and Tutz[3] kap. 4.1.1.

Definisjon 3.3.11 *Akaike's information criterion (AIC) er gitt ved*

$$AIC = -2l(\hat{\mu}) + 2(p + 1), \quad (3.21)$$

hvor p er antallet parametere i modellen vi har tilpasset.

Tilpasningen av modellen blir bedre jo flere parametere som blir benyttet. Det er allikevel en avveining om hvor mange flere parametere vi skal tillate for at tilpasningen skal bedres. AIC straffer overflødige parametere ved å legge til $2(p+1)$ til deviansen. Ved sammenligning av to modeller, vil den minste AIC'en gi den beste tilpasningen. AIC kalles ofte en straffet log-likelihood.

Kapittel 4

Resultater Gompertz

4.1 Dødelighet i befolkningen

For befolkningsdataene tar vi utgangspunkt i befolkningsstørrelsen i årene fra 1992 til 2006, og følger disse individene opp i årene fram til 2006. I en lukket befolkning ville dette vært uproblematisk, da det ikke forekommer noen emigrasjon eller immigrasjon. Vi har derimot ikke en lukket befolkning i denne oppgaven, og skal seinere vurdere implikasjonene av en slik antagelse.

Fra dataene kan vi finne den $\hat{\beta}$ og \hat{c} som optimerer likelihooden gitt i (3.14). De optimale verdiene som tilpasses dødsintensiteten for befolkningen i alderen 18 til 81 år er gitt i tabell 4.1.1. Gompertz-parameteren \hat{c} vil tilsvare til den årlige vekstfaktoren for dødsintensiteten. For kvinner øker altså dødsintensiteten årlig med 9,4%, mens økningen er noe lavere for menn på 9,1%. Til tross for en lavere vekstfaktor, har mennene en gjennomgående høyere dødsintensitet enn kvinnene.

	$\hat{\beta}$	\hat{c}
Kvinner	2.57E-05	1.0937
Menn	5.29E-05	1.0907

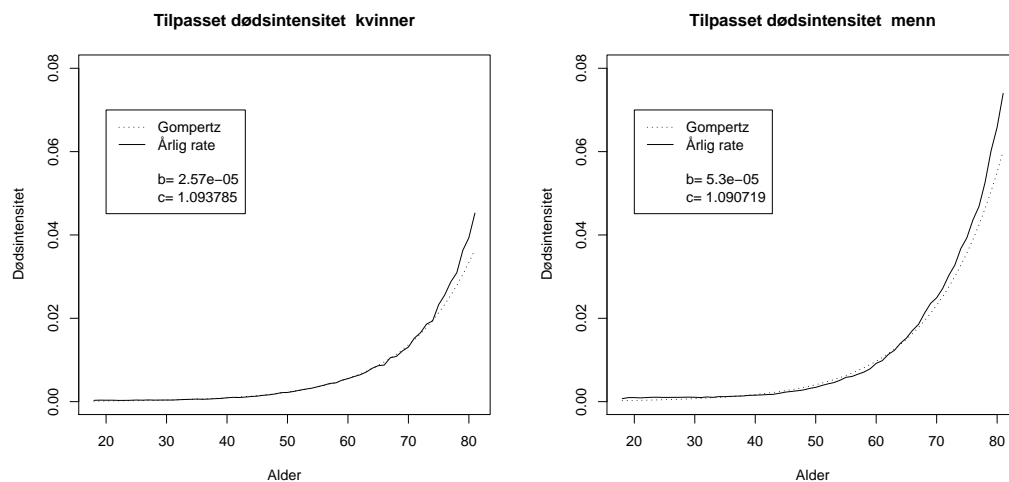
Tabell 4.1.1: *Optimale verdier for tilpasningen av Gompertz-fordelingen.*

I figur 4.1.1 ser vi hvordan dødsintensiteten for kvinner og menn utvikler seg i henhold til alder. Dette sammenlignes med de årlige ratene fra (3.8). I figur 4.1.2 vises dødsintensiteten på logaritmisk skala, og vi ser at dødsintensiteten beregnet ut ifra de årlige ratene ikke er log-lineær over hele aldersspennet.

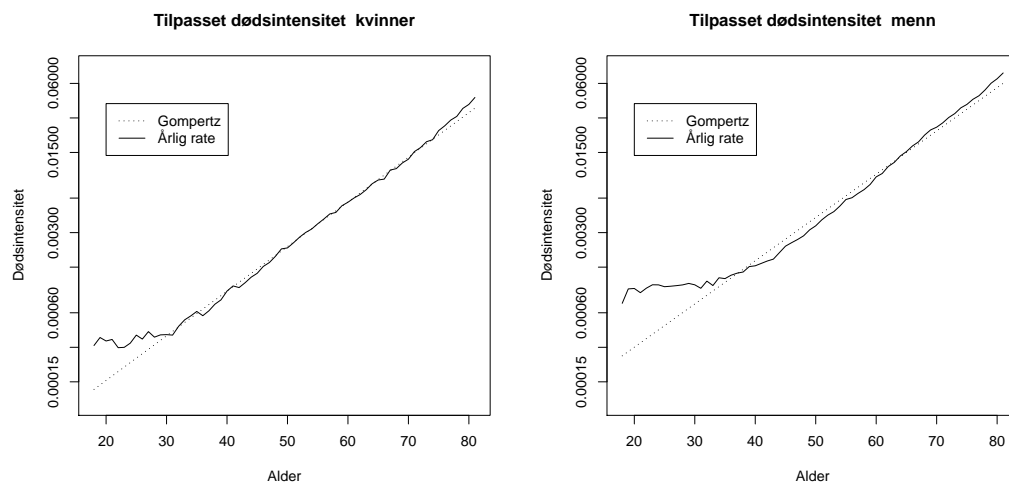
Vi observerer en høy dødsintensitet i ung alder, som holder seg tilnærmet konstant fram til 30 år for kvinner, og 35 år for menn. Deretter vokser dødsintensiteten eksponentielt for begge kjønn. Denne effekten klarer vi ikke å fange opp med Gompertz-tilpasningen. Gompertz-tilpasningen vil dermed underestimere dødsintensiteten i lav alder. I høy alder ser den også ut til en viss grad å underestimere dødsintensiteten, mens den i middels alder

vil tilsvarende overestimere dødsintensiteten. Feilestimeringen ser ut til å være størst for menn.

Det at vi foretar slike feilestimeringer av dødeligheten gir oss motivasjon til å benytte andre metoder for bedre å kunne tilpasse dødsintensiteten. For eksempel vil generaliseringen av Gompertz-fordelingen, som ble foretatt av Makeham hvor en aldersuavhengig konstant ble tillagt den eksponentielle veksten, kunne tilpasse denne effekten. Seinere i oppgaven vil også regresjonsanalyser bli benyttet til dette formålet.

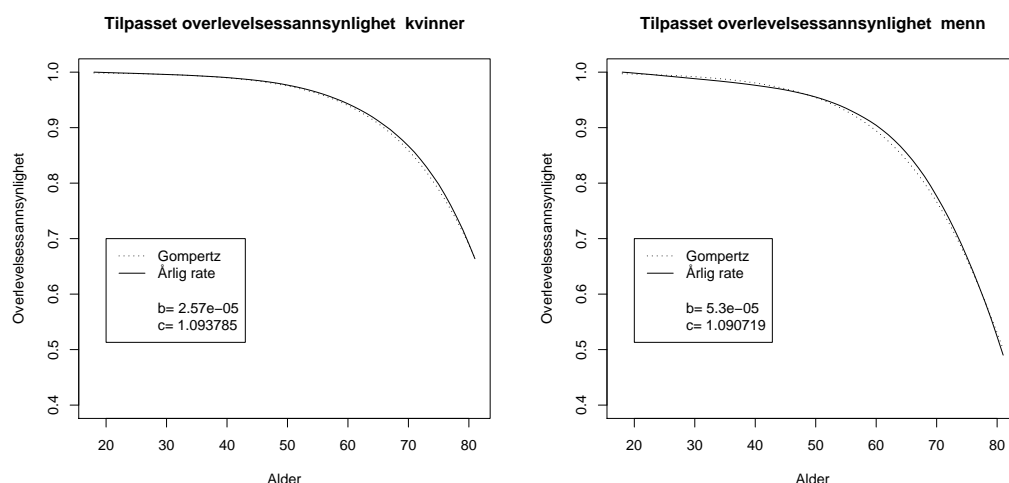


Figur 4.1.1: *Dødsintensitet for henholdsvis kvinner og menn.*



Figur 4.1.2: *Dødsintensitet for henholdsvis kvinner og menn. Logaritmisk skala.*

Ved å innsette de optimale parameterne $\hat{\beta}$ og \hat{c} inn i uttrykkene for overlevelsessannsynligheten i (3.13), kan vi sammenligne dette med ${}_t\hat{p}_x$ utledet fra de årlige ratene i (3.9). Det viser seg å være omkring 67% sannsynlighet for de 18-årige kvinnene å overleve til de er 80 år, mens sannsynligheten for menn er kun 50%. Overlevelsessannsynlighetene vises for kvinner og menn i figur 4.1.3. Konfidensintervallene gitt i (3.11) blir så små at forskjellene fra de årlige ratene ratene ikke er av interesse å framvise.



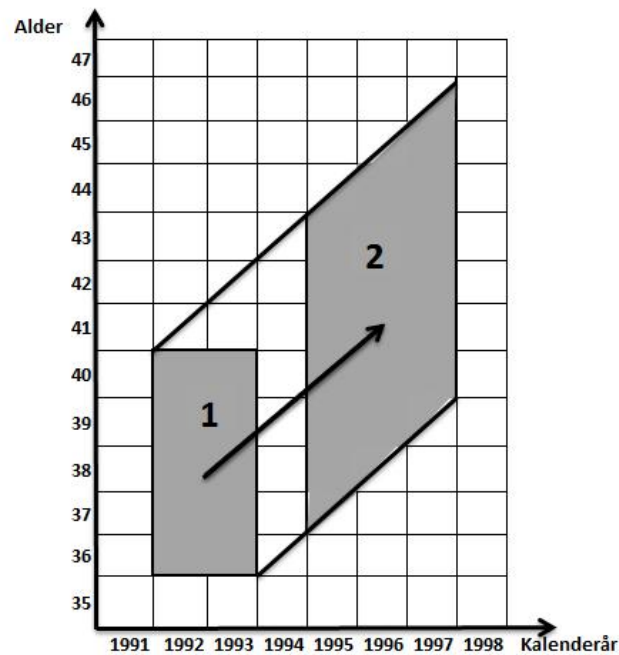
Figur 4.1.3: Overlevelsessannsynligheter fra 18 år for kvinner og menn.

4.2 Dødelighet blant uføre

For de uføre foreligger både tilgangså, alder ved tilgang og dødsår i grupper på to eller flere år. Dette skaper utfordringer i estimeringen av dødsintensitetene. Ved å tegne et Lexis-diagram for levetidsdataene synliggjøres problematikken bedre. Et eksempel på et slikt Lexis-diagram er forsøkt illustrert i figur 4.2.1. Vi får altså ikke entydige linjer, men områder hvor tilgang og død varierer innenfor. I eksempelet vises en person som er mellom 36 og 40 år og får innvilget uførestønad i årene 1992 og 1993(område 1). Personen dør en gang mellom 1995 og 1997(område 2).

Noen enkle antagelser gjøres for å konkretisere problemet, men disse skaper også usikkerhetsmomenter i dødelighetsberegningene. Vi antar en gjennomsnittsalder for personene i en aldersgruppe. Personene i alderen 36-40 år i eksempelet vil derfor få en antatt alder på 38.5 år.

En forventet observasjonstid i hvert av intervallene blir også funnet, både for de som dør i intervallet og de som overlever. Siden noen av intervallene for tilgangså delvis overlapper noen av dødsårsintervallene, vil de forventede observasjonstidene variere mellom tilgangsårene. De forventede observasjonstidene vises i tabell 4.2.1 og beregnes ut i fra simuleringer, slik det blir beskrevet i tillegg A.



Figur 4.2.1: *Lexis-diagram for levetider for de uføre. Tilgang i område 1 og død i område 2.*

Tilgangsåar	Dødsår				
	1992 – 1994	1995 – 1997	1998 – 2000	2001 – 2003	2004 – 2006
1992 – 1993	1.000/2.000	1.500/3.000	1.500/3.000	1.500/3.000	1.500/3.000
1994 – 1995	0.250/0.500	1.625/3.000	1.500/3.000	1.500/3.000	1.500/3.000
1996 – 1997		0.500/1.000	1.500/3.000	1.500/3.000	1.500/3.000
1998 – 1999			1.000/2.000	1.500/3.000	1.500/3.000
2000 – 2001			0.250/0.500	1.625/3.000	1.500/3.000
2002 – 2003				0.500/1.000	1.500/3.000
2004 – 2005					1.000/2.000

Tabell 4.2.1: *Anslåtte observasjonstider i kohortene for uføre. Gitt død i intervallet/gitt overlevelse.*

De optimale parameterne for tilpasningen av uføredødeligheten til en Gompertz-fordeling vises i tabell 4.2.2. For befolkningen estimerte vi den årlige veksten i dødsintensiteten til å være i overkant av 9%. Veksten vi estimerer for de uføre er betraktelig lavere for samtlige uføreårsaker. Lavest er den årlige veksten i dødsintensiteten for gruppa uførepensjonert grunnet andre årsaker, hvor den er under 3%. Imidlertid oppnår vi høyere $\hat{\beta}$ -verdier for de uføre enn for befolkningen. Dette gjør at de uføre har høye dødsintensiteter i ung alder.

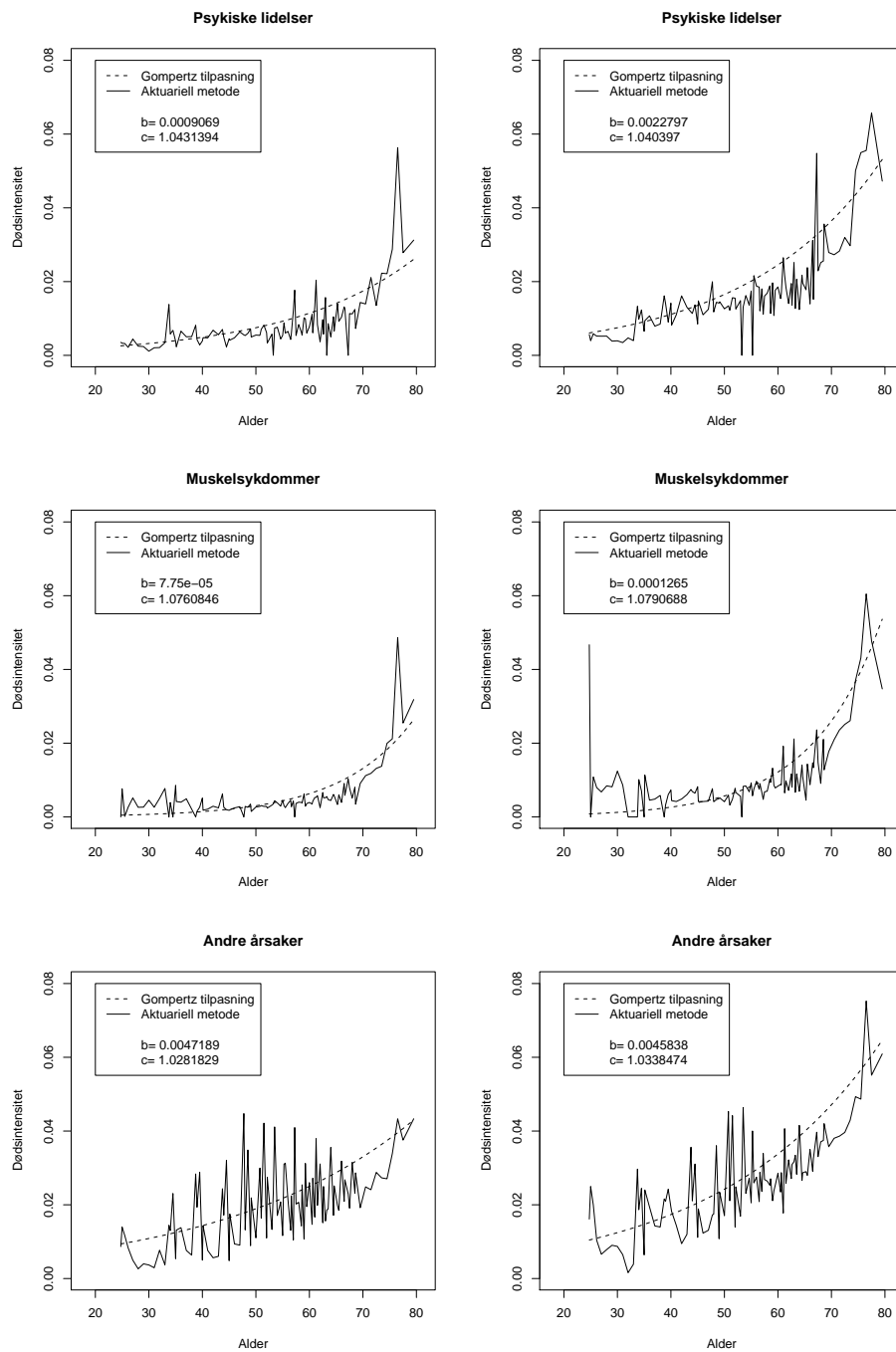
	Årsak	$\hat{\beta}$	\hat{c}
Kvinner	Psykiske lidelser	9.07E-04	1.0431
	Muskelsykdommer	7.75E-05	1.0761
	Andre årsaker	4.72E-03	1.0282
Menn	Psykiske lidelser	2.27E-03	1.0404
	Muskelsykdommer	1.26E-04	1.0791
	Andre årsaker	4.58E-03	1.0338

Tabell 4.2.2: *Optimale verdier for tilpasningen av Gompertz-fordelingen.*

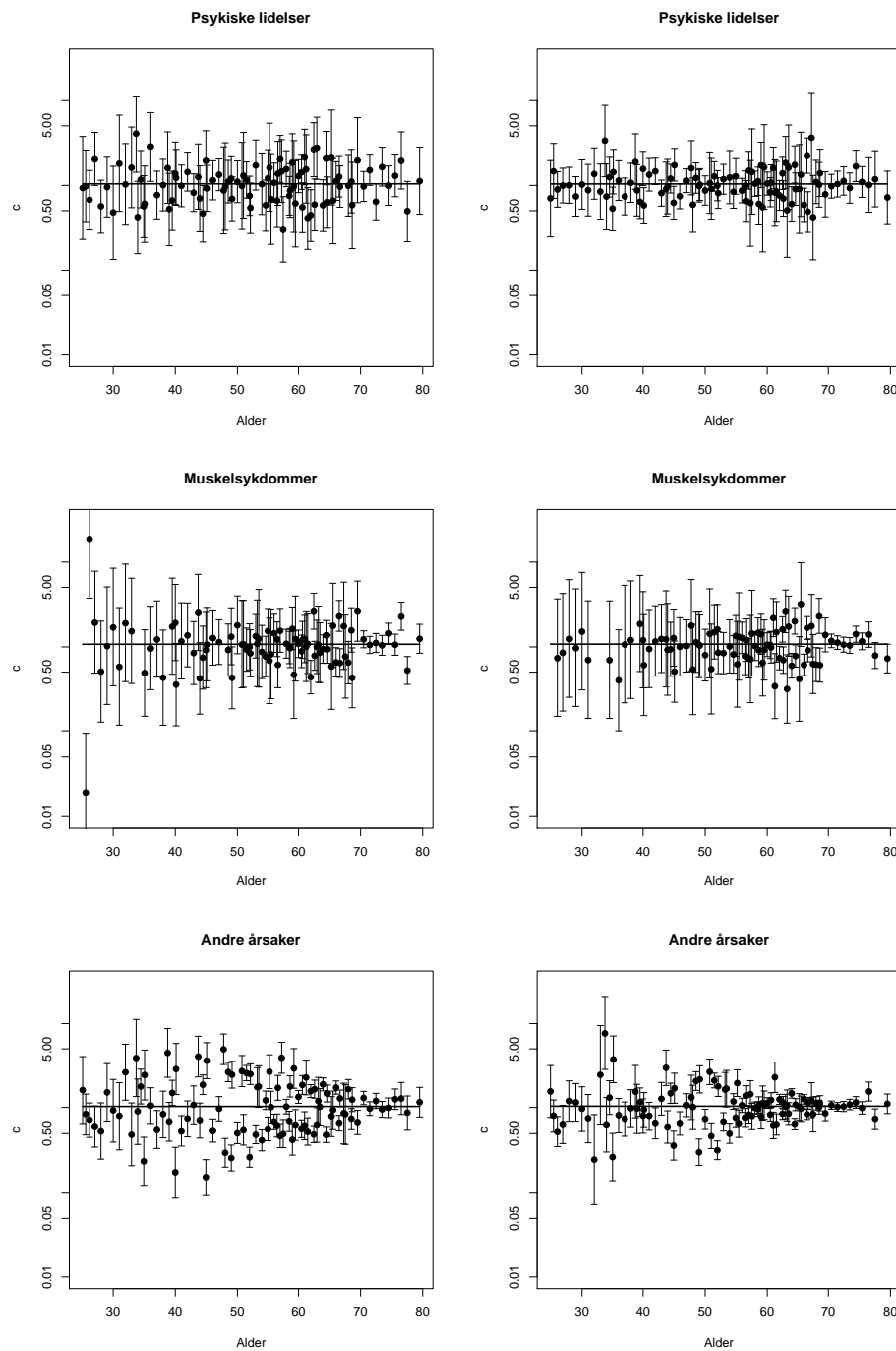
Det er store svingninger i dødsintensiteten fra de årlige ratene, slik figur 4.2.2 viser. Dette er fordi vi har få uføre og kun små endringer i antallet døde gir store utslag i ratene. Kraftigst er svingningene for gruppa uførepensjonert av andre årsaker. For gruppene uførepensjonert grunnet psykiske lidelser og muskelsykdommer, ser dødsintensiteten ut til å holde seg stabil, men noe økende for menn. Dette gjelder fram til de uførepensjonerte er rundt 70 år og dødsintensiteten øker kraftig. Gompertz-fordelingen klarer ikke helt å fange opp denne økningen i vekstraten, noe som gjør at dødsintensiteten blir underestimert i høy alder. Dødsintensitetene blir overestimert i 60-70 års alderen. Det viser seg altså at vi ikke har en log-lineær vekst i dødsintensiteten over hele aldersspennet. Tilpasningsproblematikken tilsvarende den vi hadde for befolkningsdataene.

For å sjekke tilpasningen av parametrene i Gompertz-modellen, vises konfidensintervallene for sannsynlighetsmaksimerings-resultatet til vekstratene som ble utledet i (3.16). Ved å sjekke om den tilpassede vekstraten, \hat{c} , er inneholdt i konfidensintervallene, kan vi si litt om hvor god tilpasningen er. Dette vises i figur 4.2.3 for henholdsvis menn og kvinner. Konfidensintervallene kan se ut til å være størst i ung alder. Dette kommer av at vi i beregningen av konfidensintervaller deler på antallet døde i intervallet, og i ung alder forekommer det få dødsfall. For de uførepensjonerte grunnet andre årsaker, er konfidensintervallene gjennomgående mindre enn for de andre uføreårsakene over hele aldersperspektivet. Tilpasningen er inneholdt i få av konfidensintervallene. Det ser allikevel ut til at vi ikke foretar noen systematisk feil i tilpasningen, men at konfidensintervallene er tilfeldig spredt rundt tilpasningsverdien.

I noen av de yngre aldersgruppene, spesielt for gruppa uførepensjonert grunnet muskelsykdommer, observeres det ingen dødsfall. I vår interpretasjon av vekstraten, vil dette gjøre at vi dividerer med null. Disse er derfor utelatt fra figuren. Utenom dette er den tilpassede c -verdien stort sett inneholdt i konfidensintervallene vi estimerer, og tilpasningen er tilfredsstillende.



Figur 4.2.2: Dødsintensiteter for de uførepensjonerte. Årlige rater og tilpasset Gompertz-fordeling. Kvinner til venstre og menn til høyre.



Figur 4.2.3: Konfidensintervaller for parametertilpasningen av c i Gompertz-fordelingen for uførepensjonerte. Kvinner til venstre og menn til høyre. Logaritmisk skala.

Kapittel 5

Resultater regresjon

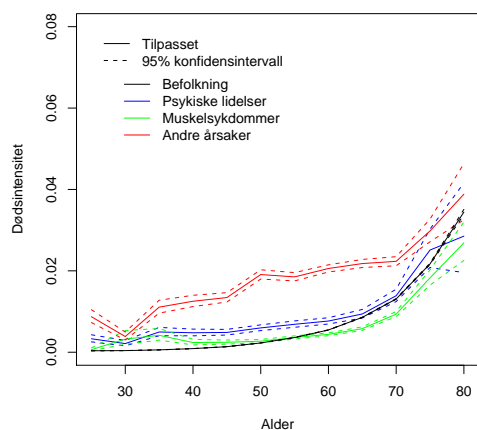
5.1 Effekt av alder ved død

Vi vurderer først utviklingen av dødsintensiteten kun som følge av alder ved død. Til dette benyttes modellen ADØD med dødsintensiteten $\mu_a = \exp(\beta_0 + \beta_a)$ for en alder a . Dødsintensiteten vises i figur 5.1.1 for de ulike uføregruppene og befolkningen med et tilnærmet 95% konfidensintervall. Dette tilsvarer spesialtilfellet fra eksempel 3.3.5, hvor kurvene gattes ved å se på aldersintervaller på 5 år.

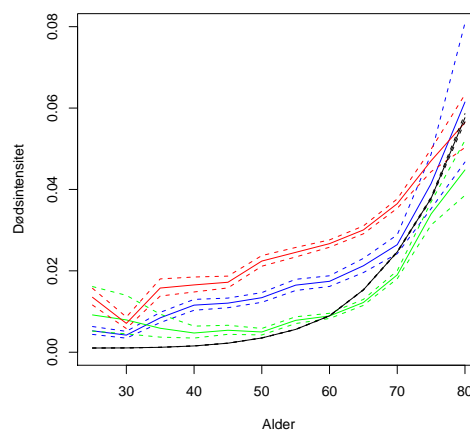
Vi ser at de uføre i hovedsak har en høyere dødsintensitet enn befolkningen, og dødsintensiteten er sterkt aldersbetinget. Det er de uførepensjonerte grunnet andre årsaker som har den høyeste dødsintensiteten hos begge kjønn. Dødsintensiteten til de uførepensjonerte grunnet psykiske lidelser har samme aldersutvikling som for de grunnet andre årsaker, men på et noe lavere nivå. I gruppa av uførepensjonerte grunnet muskelsykdommer er dødsintensiteten blant de yngste kvinnene lav, men sterkt voksende de første årene. Deretter avtar den noe fram til de uføre når 40 år. De muskelsyke mennene har en avtagende dødsintensitet fram til de når 50 år. På grunn av få dødsfall i ung alder ser vi også at konfidensintervallene vi beregner for denne gruppa blir ganske vide. Dette gjør at estimatene her er usikre.

Ved å se på dødsintensiteten på log-skala i figur 5.1.1(c) og 5.1.1(d), synliggjøres forskjellene mellom dødsintensiteten i ung alder. For gruppa uførepensjonert grunnet psykiske lidelser og for gruppa med andre årsaker har vi, fra alder omkring 35 år, en eksponentiell vekst i dødsintensiteten. Den eksponentielle veksten inntreffer fra omkring 40 år for de med muskelsykdommer. Det betyr betyr at Gompertz-antagelsene er plausible fra disse alderne.

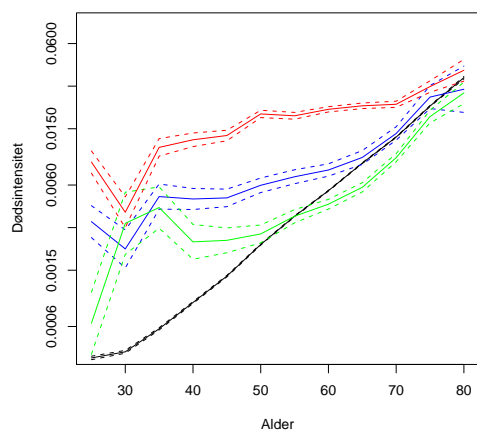
Ettersom de uføre eldes, går uføredødelighetene mot befolkningsdødeligheten. Uføredatene inngår som kjent i befolkningsdataene. Når de uføre eldes, og flere uførepensjoneres, øker også andelen uføre i befolkningsdataene. Det er derfor naturlig at dødelighetene går mot hverandre.



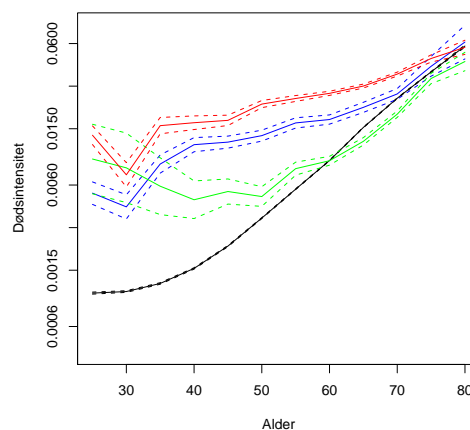
5.1.1(a): Dødsintensitet kvinner



5.1.1(b): Dødsintensitet menn



5.1.1(c): Dødsintensitet kvinner, log-skala



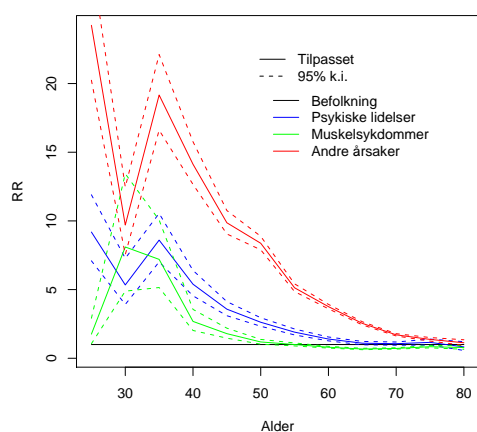
5.1.1(d): Dødsintensitet menn, log-skala

Figur 5.1.1: *Dødsintensiteter for kvinner og menn fra 20 år, ikke tatt hensyn til tilgangår. Modell ADØD.*

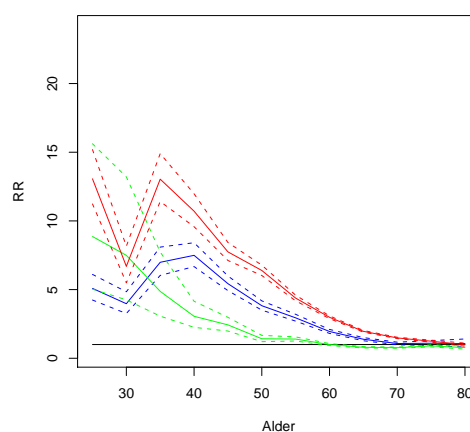
For å kvantifisere forskjellene fra befolkningsdødeligheten ser vi på risiko for død i uføregruppene relativt til den generelle befolkningen, slik beskrevet i proposisjon 3.3.7. Den relative risikoen vises i figur 5.1.2, og listes i tabellene 5.1.1 og 5.1.2.

Den relative risikoen er størst i ung alder. Dødsintensiteten er opp imot 25 ganger så høy blant de kvinnelig uføre grunnet andre årsaker og 15 ganger så høy for den samme gruppa av menn som for befolkningen. For de uførepensjonerte grunnet psykiske lidelser og gruppa uførepensjonerte grunnet muskelsykdommer er dødsintensiteten rundt 5-10 ganger så høy som for befolkningen i ung alder. Figuren viser også at reduksjonen i den relative risikoen

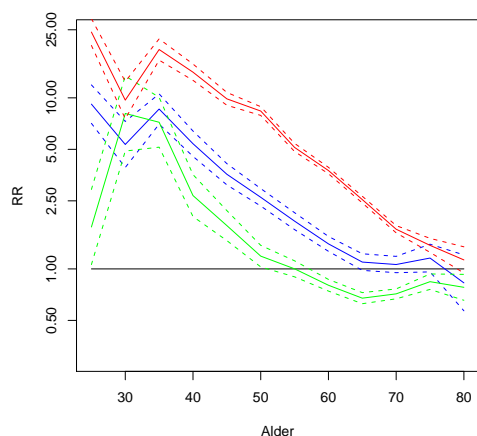
blant alle uføregruppene er tilnærmet log-lineær fra alder omkring 35 år for begge kjønn. Overdødeligheten synker kraftig blant gruppa uførepensjonert grunnet muskelsykdommer. Denne gruppa har, i visse aldre, en signifikant lavere dødsintensitet enn befolkningen. For kvinnene er den relative risikoen nede i $0.67(0.63-0.73)$ og $0.77(0.73-0.81)$ for menn. Et tilnærmet 95% konfidensintervall vises i parentes. Dødsintensiteten er signifikant lavere enn befolkningens fra kvinnene er 55 år og for mennene fra de er omkring 60 år. For de uføretrygdde grunnet psykiske lidelser kan vi ikke påvise signifikante forskjeller i dødsintensiteten fra befolknigen.



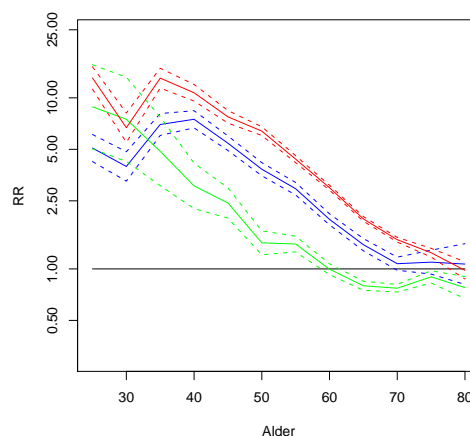
5.1.2(a): Relativ risiko kvinner



5.1.2(b): Relativ risiko menn



5.1.2(c): Relativ risiko kvinner, log-skala



5.1.2(d): Relativ risiko menn, log-skala

Figur 5.1.2: Relativ risiko for død i uføregruppene relativt til befolkningsdødeligheten i modellen ADØD. Naturlig og logaritmisk skala.

	P	M	A
Alder	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
17.5-22.5	9.19 (7.09 - 11.90)	1.76 (1.06 - 2.91)	24.24 (20.24 - 29.03)
22.5-27.5	5.34 (3.92 - 7.28)	8.10 (4.89 - 13.43)	9.70 (7.55 - 12.46)
27.5-32.5	8.60 (7.02 - 10.55)	7.20 (5.15 - 10.07)	19.16 (16.61 - 22.11)
32.5-37.5	5.39 (4.55 - 6.40)	2.68 (2.02 - 3.56)	14.13 (12.66 - 15.76)
37.5-42.5	3.58 (3.10 - 4.13)	1.79 (1.46 - 2.19)	9.87 (9.05 - 10.75)
42.5-47.5	2.62 (2.33 - 2.95)	1.19 (1.03 - 1.37)	8.37 (7.89 - 8.88)
52.5-57.5	1.90 (1.70 - 2.13)	1.00 (0.90 - 1.11)	5.13 (4.86 - 5.42)
57.5-62.5	1.40 (1.27 - 1.55)	0.80 (0.74 - 0.87)	3.76 (3.60 - 3.92)
62.5-67.5	1.10 (0.98 - 1.23)	0.67 (0.63 - 0.73)	2.54 (2.43 - 2.66)
67.5-72.5	1.06 (0.95 - 1.18)	0.71 (0.67 - 0.76)	1.71 (1.62 - 1.79)
72.5-77.5	1.16 (0.96 - 1.39)	0.84 (0.76 - 0.93)	1.37 (1.25 - 1.51)
77.5-82.5	0.83 (0.57 - 1.21)	0.78 (0.66 - 0.93)	1.13 (0.95 - 1.35)

Tabell 5.1.1: *Relativ risiko for død i uføregruppene relativt til befolkningsdødeligheten i modellen ADØD. Kvinner.*

	P	M	A
Alder	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
17.5-22.5	5.09 (4.24 - 6.11)	8.86 (5.03 - 15.62)	13.07 (11.24 - 15.19)
22.5-27.5	3.97 (3.26 - 4.84)	7.49 (4.25 - 13.20)	6.69 (5.49 - 8.16)
27.5-32.5	6.99 (6.03 - 8.09)	4.87 (3.07 - 7.73)	13.04 (11.41 - 14.90)
32.5-37.5	7.49 (6.67 - 8.41)	3.06 (2.25 - 4.15)	10.71 (9.58 - 11.97)
37.5-42.5	5.43 (4.92 - 5.99)	2.43 (1.98 - 2.98)	7.74 (7.12 - 8.43)
42.5-47.5	3.83 (3.50 - 4.18)	1.42 (1.21 - 1.67)	6.39 (6.03 - 6.78)
52.5-57.5	2.94 (2.70 - 3.20)	1.40 (1.26 - 1.55)	4.38 (4.18 - 4.59)
57.5-62.5	1.95 (1.81 - 2.10)	1.00 (0.93 - 1.07)	2.98 (2.89 - 3.08)
62.5-67.5	1.39 (1.28 - 1.50)	0.80 (0.75 - 0.85)	1.96 (1.90 - 2.02)
67.5-72.5	1.07 (0.98 - 1.17)	0.77 (0.73 - 0.81)	1.48 (1.44 - 1.53)
72.5-77.5	1.10 (0.93 - 1.29)	0.90 (0.83 - 0.98)	1.24 (1.17 - 1.32)
77.5-82.5	1.07 (0.81 - 1.40)	0.78 (0.67 - 0.90)	0.98 (0.87 - 1.10)

Tabell 5.1.2: *Relativ risiko for død i uføregruppene relativt til befolkningsdødeligheten i modellen ADØD. Menn.*

5.2 Effekt av tilgangssår

Modellen utvides nå med forklaringsvariabelen tilgangssår til $ADØD + TILGANG$. Dette gir dødsintensiteten $\mu_{a,t} = \exp(\beta_0 + \beta_a + \beta_t)$ for alder a ved død og tilgang i år t .

For befolkningen benyttes året da oppfølgingen startet som en forklaringsvariabel tilsvarende tilgangssår for de uføre. Dermed kan vi sammenligne effekten av året da uførepensjoneringen skjer mot den generelle endringen i befolkningsdødeligheten.

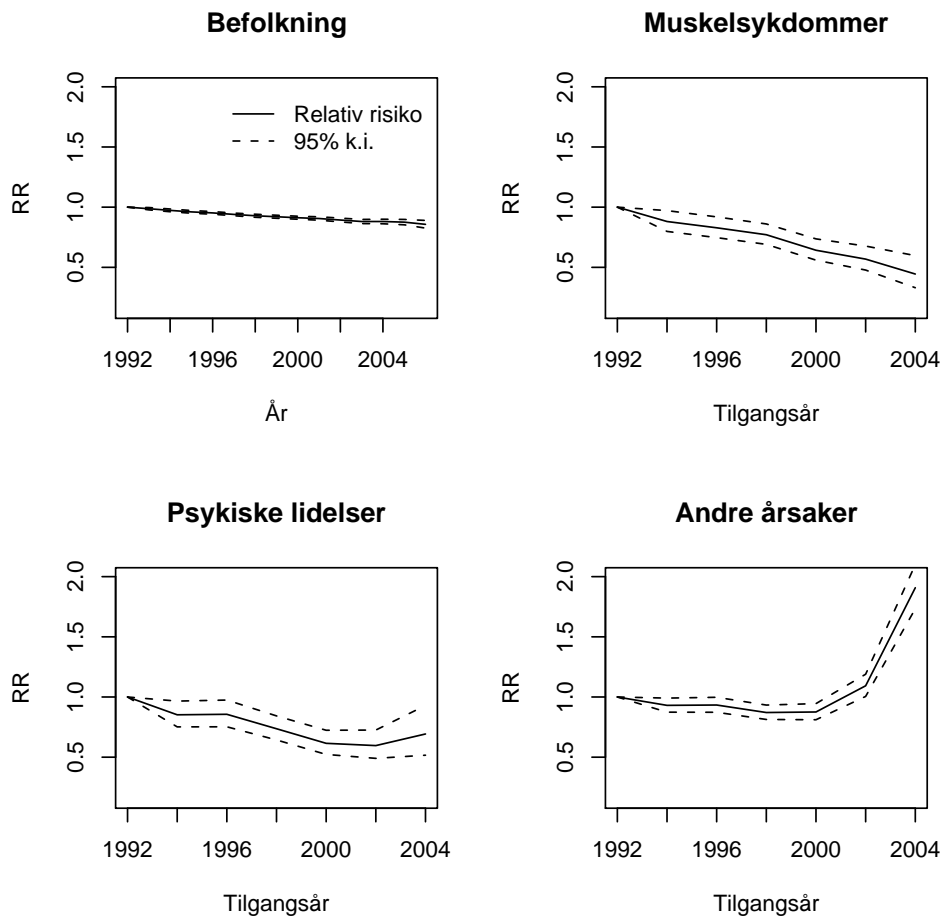
Den relative risikoen for død i ulike tilgangssår relativt til død ved tilgang i 1992-1993, vises i figurene 5.2.1 og 5.2.2, og listes i tabellene 5.2.1 og 5.2.2. Vi ser tydelige reduksjoner i dødsintensiteten ettersom tilgangsåret øker. Alle uføreårsakene har en signifikant kraftigere reduksjon i dødsintensiteten enn den generelle befolkningen.

Vi kan anta en lineær reduksjon i dødsintensiteten for befolkningen, de med psykiske lidelser og de med muskelsykdommer. For andre årsaker er reduksjonen lineær fram til tilgang i 2002-2003, hvor dødsintensiteten øker kraftig for begge kjønn. Effekten synes å være størst for kvinner. En slik økning i dødsintensiteten, kan muligens forklares av endringen i regelverket for tildeling av uførepensjon. Vi kan tenke oss at de uføre som etter nytt reglement 1.1.2004 gis rett til tidsbegrenset uførestønad, tidligere har bidratt med en lavere dødsintensitet enn resten av uførepensjonistene. Den gjenværende bestanden av varige uføre vil dermed få en høyere dødsintensitet. Det er tydelig at denne effekten er størst blant de som uførepensjoneres av andre årsaker, men vi ser svake økninger i dødsintensiteten mot siste tilgangssår for gruppa av menn uførepensjonert grunnet psykiske lidelser og kvinner uførepensjonert grunnet muskelsykdommer.

I figur 5.2.3 ser vi hvordan dødsintensiteten for 50-årige kvinner og menn endres over ulike tilgangssår. Dødsintensiteten for de uførepensjonerte grunnet muskelsykdommer tilsvarer nesten befolkningens for begge kjønn. Dødsintensiteten for de med psykiske lidelser ligger noe høyere, spesielt for menn. De uførepensjonerte grunnet andre årsaker har en betraktelig høyere dødsintensitet.

Vi beregner deviansen gitt fra definisjon 3.3.9. Dermed kan vi vurdere hvorvidt endringen i deviansen fra modeller med færre forklaringsvariabler er signifikante. Endringene vises i tabellene 5.2.3 og 5.2.4. Vi ser at faktoren $ADØD$ i høyeste grad er signifikant i modellforklaringen. For begge kjønn ser vi også en signifikant endring i deviansen ved å legge til forklaringsvariabelen $TILGANG$ i modellen. Endringen i deviansen er størst for gruppa uførepensjonert grunnet andre årsaker. Denne gruppa har også den største deviansen fra den mettede modellen.

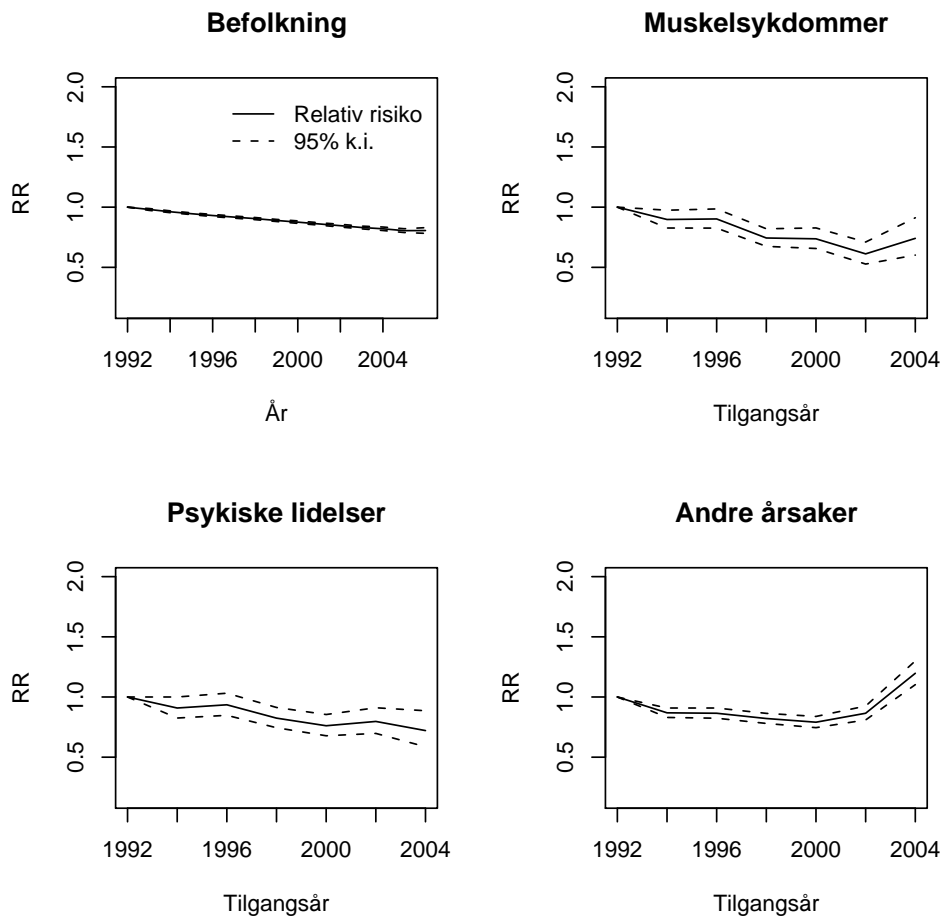
Interaksjonen mellom alder ved død og tilgangssår tas så hensyn til med modellen $ADØD * TILGANG$. Blant de kvinnelige uførepensjonerte grunnet psykiske lidelser kan vi ikke påvise noen interaksjon mellom alder ved død og tilgangssår. Med unntak av disse, er endringen i deviansen signifikant når vi tilpasser et interaksjonsledd i modellen.



Figur 5.2.1: *Relativ risiko for død i ulike tilgangssår relativt til død i første tilgangssår. Modell ADØD + TILGANG. Kvinner.*

	B	P	M	A
Tilgang	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
92-93	1.00	1.00	1.00	1.00
94-95	0.97(0.96-0.98)	0.85(0.75-0.97)	0.88(0.80-0.97)	0.93(0.87-0.99)
96-97	0.95(0.94-0.96)	0.86(0.75-0.97)	0.83(0.75-0.92)	0.93(0.87-1.00)
98-99	0.93(0.92-0.94)	0.74(0.64-0.84)	0.77(0.69-0.86)	0.87(0.81-0.93)
00-01	0.91(0.90-0.92)	0.62(0.52-0.72)	0.64(0.56-0.74)	0.88(0.81-0.95)
02-03	0.89(0.88-0.91)	0.60(0.49-0.73)	0.57(0.48-0.68)	1.09(1.00-1.19)
04-05	0.88(0.86-0.90)	0.69(0.52-0.93)	0.44(0.33-0.60)	1.91(1.73-2.10)

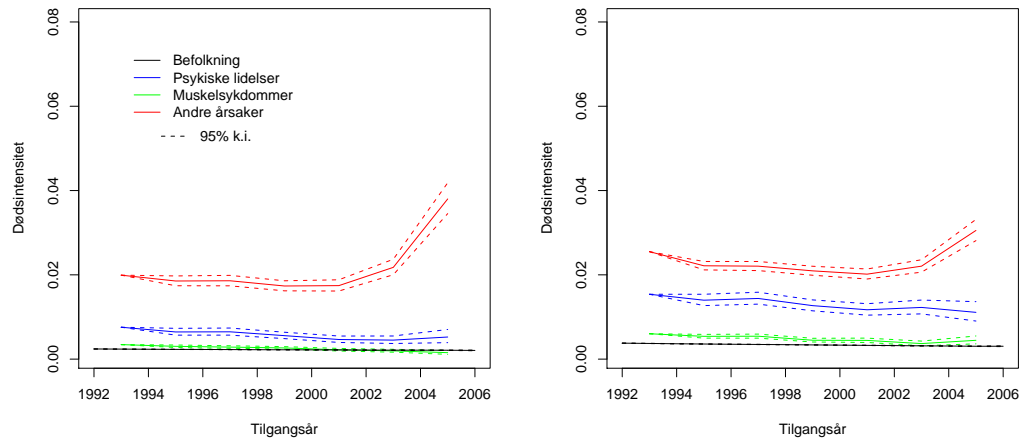
Tabell 5.2.1: *Relativ risiko for død i ulike tilgangssår relativt til død i første tilgangssår. Modell ADØD + TILGANG. Kvinner.*



Figur 5.2.2: *Relativ risiko for død i ulike tilgangår relativt til død i første tilgangår. Modell ADØD + TILGANG. Menn.*

	B	P	M	A
Tilgang	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
92-93	1.00	1.00	1.00	1.00
94-95	0.96(0.95-0.97)	0.91(0.83-1.00)	0.90(0.83-0.97)	0.87(0.83-0.91)
96-97	0.93(0.92-0.94)	0.94(0.85-1.03)	0.90(0.83-0.99)	0.87(0.82-0.91)
98-99	0.90(0.90-0.91)	0.82(0.74-0.91)	0.74(0.67-0.82)	0.82(0.78-0.86)
00-01	0.88(0.87-0.89)	0.76(0.68-0.85)	0.74(0.66-0.83)	0.79(0.75-0.84)
02-03	0.85(0.83-0.86)	0.80(0.70-0.91)	0.61(0.53-0.71)	0.87(0.81-0.92)
04-05	0.82(0.81-0.83)	0.72(0.59-0.89)	0.74(0.60-0.91)	1.20(1.10-1.30)

Tabell 5.2.2: *Relativ risiko for død i ulike tilgangår relativt til død i første tilgangår. Modell ADØD + TILGANG. Menn.*



Figur 5.2.3: Dødsintensitet for henholdsvis en 50-årig kvinne og mann for ulike tilgangår. Modell ADØD + TILGANG.

Årsak	Modell	Res.df	Res.dev	df	Endr.dev	p-verdi
P	1	321	981.88			
	+ADØD	310	475.80	11	506.08	<0.0001
	+TILGANG	304	419.88	6	55.91	<0.0001
	+ADØD*TILGANG	248	345.76	56	74.13	0.05
M	1	321	1894.12			
	+ADØD	310	599.26	11	1294.86	<0.0001
	+TILGANG	304	508.12	6	91.15	<0.0001
	+ADØD*TILGANG	248	380.26	56	127.86	<0.0001
A	1	321	3143.38			
	+ADØD	310	2351.79	11	791.59	<0.0001
	+TILGANG	304	2097.73	6	254.06	<0.0001
	+ADØD*TILGANG	248	1942.42	56	155.31	<0.0001

Tabell 5.2.3: Deviansanalyser kvinner.

Årsak	Modell	Res.df	Res.dev	df	Endr.dev	p-verdi
P	1	321	1402.15			
	+ADØD	310	455.17	11	946.98	<0.0001
	+TILGANG	304	419.76	6	35.41	<0.0001
	+ADØD*TILGANG	248	329.98	56	89.78	0.0028
M	1	321	1957.70			
	+ADØD	310	615.84	11	1341.86	<0.0001
	+TILGANG	304	539.05	6	76.79	<0.0001
	+ADØD*TILGANG	248	454.54	56	84.51	0.01
A	1	321	2950.38			
	+ADØD	310	1449.28	11	1501.10	<0.0001
	+TILGANG	304	1299.04	6	150.24	<0.0001
	+ADØD*TILGANG	248	1120.92	56	178.12	<0.0001

Tabell 5.2.4: *Deviansanalyser menn.*

5.3 Effekt av tid fra tilgang

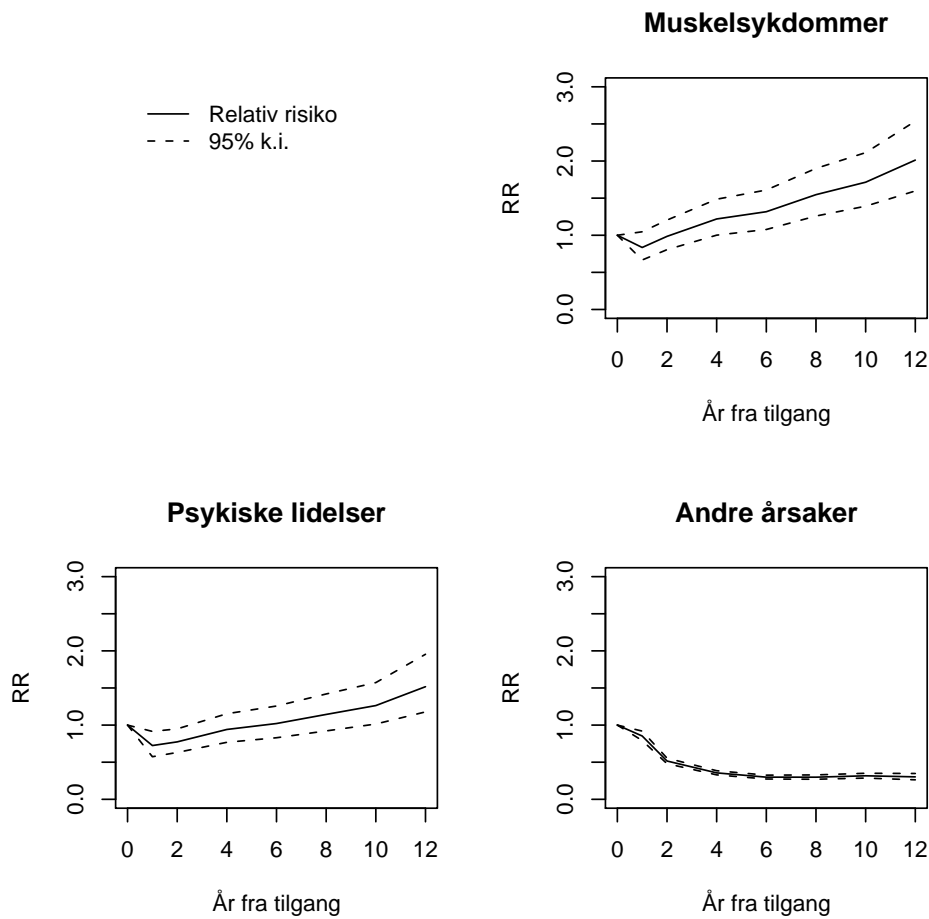
Med regresjonsmodellen ADØD + TID FRA TILG ser vi på hvordan dødsintensiteten endres ettersom tiden fra tilgang øker. Dette gir dødsintensiteten $\mu_{a,f} = \exp(\beta_0 + \beta_a + \beta_f)$ for alder ved død a og år fra tilgang f .

Relativ risiko for død etter ulik tid fra tilgang relativt til død i året for tilgang, vises i figurene 5.3.1 og 5.3.2, og listes i tabellene 5.3.1 og 5.3.2.

For kvinner uførepensjonert grunnet muskelsykdommer og psykiske lidelser, reduseres dødsintensiteten signifikant det første året etter uførepensjonering. De påfølgende årene øker den jevnt. Spesielt kraftig er økningen hos de uførepensjonerte grunnet muskelsykdommer. Blant mennene ser vi ikke en signifikant reduksjon i dødsintensiteten det første året etter uførepensjonering, verken blant de muskelsyke eller for de med psykiske lidelser.

For uførepensjonerte grunnet andre årsaker, er dødsintensiteten signifikant avtagende de to første årene etter tilgang, og holder seg lavere gjennom hele oppfølgingsperioden. Reduksjonen er kraftigst blant kvinnene. For andre årsaker er usikkerheten i den relative risikoen liten. Konfidensintervallene til de relative risikoene for de uførepensjonerte grunnet muskelsykdommer og psykiske lidelser er betraktelig videre.

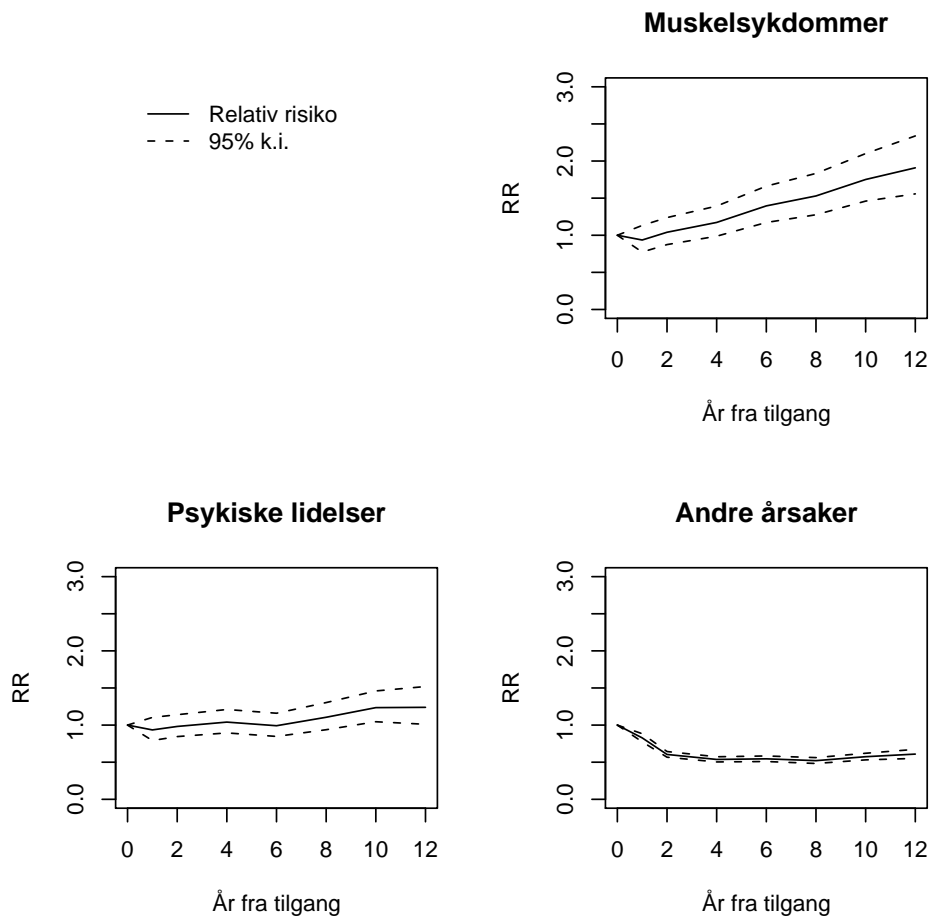
I figur 5.3.3 vises dødsintensiteten for en 50 år gammel person for ulike år fra tilgang. Dødsintensiteten blant de som uførepensjoneres av andre årsaker er spesielt høy de første årene etter tilgang, men stabiliserer seg etterhvert med samme vekstrate som de resterende uføreårsakene. Umiddelbar dødsrisiko etter tilgang til uførepensjon er, for kvinnene i gruppa med andre årsaker, høyere enn for den samme gruppa av menn. Dødsintensiteten er ellers gjennomgående lavere for kvinner enn for menn.



Figur 5.3.1: *Relativ risiko for død for ulik tid fra tilgang relativt til død i tilgangsåret. Modell ADØD + TID FRA TILGANG. Kvinner.*

	P	M	A
Tid fra tilgang	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
0	1.00	1.00	1.00
1	0.72(0.57-0.91)	0.84(0.67-1.05)	0.85(0.79-0.92)
2	0.77(0.63-0.95)	0.99(0.81-1.20)	0.52(0.48-0.55)
4	0.94(0.77-1.15)	1.22(1.00-1.49)	0.36(0.33-0.39)
6	1.02(0.83-1.26)	1.32(1.08-1.61)	0.30(0.27-0.33)
8	1.14(0.92-1.42)	1.55(1.26-1.90)	0.30(0.27-0.33)
10	1.26(1.01-1.57)	1.71(1.39-2.11)	0.32(0.29-0.35)
12	1.52(1.18-1.95)	2.01(1.60-2.54)	0.30(0.26-0.35)

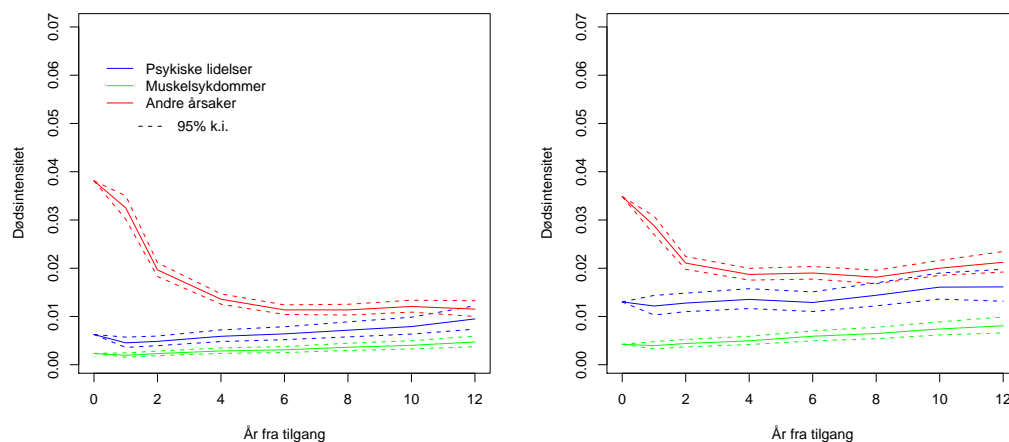
Tabell 5.3.1: *Relativ risiko for død for ulik tid fra tilgang relativt til død i tilgangsåret. Modell ADØD + TID FRA TILGANG. Kvinner.*



Figur 5.3.2: *Relativ risiko for død for ulik tid fra tilgang relativt til død i tilgangsåret. Modell ADØD + TID FRA TILGANG. Menn.*

	P	M	A
Tid fra tilgang	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
0	1.00	1.00	1.00
1	0.93(0.79-1.10)	0.94(0.77-1.13)	0.83(0.77-0.88)
2	0.98(0.85-1.14)	1.04(0.87-1.24)	0.60(0.57-0.64)
4	1.04(0.89-1.21)	1.17(0.99-1.39)	0.54(0.50-0.57)
6	0.99(0.85-1.16)	1.40(1.17-1.66)	0.55(0.51-0.58)
8	1.11(0.94-1.30)	1.53(1.28-1.83)	0.52(0.48-0.56)
10	1.24(1.05-1.46)	1.75(1.46-2.10)	0.57(0.53-0.62)
12	1.24(1.01-1.52)	1.91(1.56-2.34)	0.61(0.55-0.67)

Tabell 5.3.2: *Relativ risiko for død for ulik tid fra tilgang relativt til død i tilgangsåret. Modell ADØD + TID FRA TILGANG. Menn.*



Figur 5.3.3: Dødsintensitet for henholdsvis en 50-årig kvinne og 50-årig mann for ulik tid fra tilgang. Modell ADØD + TID FRA TILGANG.

Resultatene fra deviansanalysene med modellen ADØD*TID FRA TILGANG listes i tabellene 5.3.3 og 5.3.4. Fra disse resultatene ser vi at TID FRA TILGANG bidrar sterkt til forklaring av modellen. Endringen vi får i deviansen ved å inkludere tid fra tilgang som en forklaringsvariabel er signifikant for samtlige av uforeårsakene. Deviansendringen er spesielt stor for gruppa uførepensjonert grunnet andre årsaker, og da særlig for kvinner.

Interaksjon mellom de to forklaringsvariablene for de uførepensjonerte kvinnene grunnet psykiske lidelser kan ikke påvises. Effekten ved å tilpasse et interaksjonsledd er heller ikke veldig stor for menn med muskelsykdommer. Interaksjonen er derimot signifikant for de resterende årsakene.

Årsak	Modell	Res.df	Res.dev	df	Endr.dev	p-verdi
P	1	321	981.88			
	+ADØD	310	475.80	11	506.08	<0.0001
	+TID FRA TILGANG	303	403.28	7	72.52	<0.0001
	+ADØD*TID FRA TILGANG	249	330.85	54	72.43	0.05
M	1	321	1894.12			
	+ADØD	310	599.26	11	1294.86	<0.0001
	+TID FRA TILGANG	303	463.78	7	135.48	<0.0001
	+ADØD*TID FRA TILGANG	249	360.61	54	103.17	<0.0001
A	1	321	3143.38			
	+ADØD	310	2351.79	11	791.59	<0.0001
	+TID FRA TILGANG	303	810.83	7	1540.96	<0.0001
	+ADØD*TID FRA TILGANG	249	537.88	54	272.95	<0.0001

Tabell 5.3.3: Deviansanalyser kvinner.

Årsak	Modell	Res.df	Res.dev	df	Endr.dev	p-verdi
P	1	321	1402.15			
	+ADØD	310	455.17	11	946.98	<0.0001
	+TID FRA TILGANG	303	430.94	7	24.23	0.001
	+ADØD*TID FRA TILGANG	249	321.02	54	109.92	<0.0001
M	1	321	1957.70			<0.0001
	+ADØD	310	615.84	11	1341.86	<0.0001
	+TID FRA TILGANG	303	473.34	7	142.50	<0.0001
	+ADØD*TID FRA TILGANG	249	394.84	54	78.50	0.02
A	1	321	2950.38			<0.0001
	+ADØD	310	1449.28	11	1501.10	<0.0001
	+TID FRA TILGANG	303	886.97	7	562.30	<0.0001
	+ADØD*TID FRA TILGANG	249	598.72	54	288.26	<0.0001

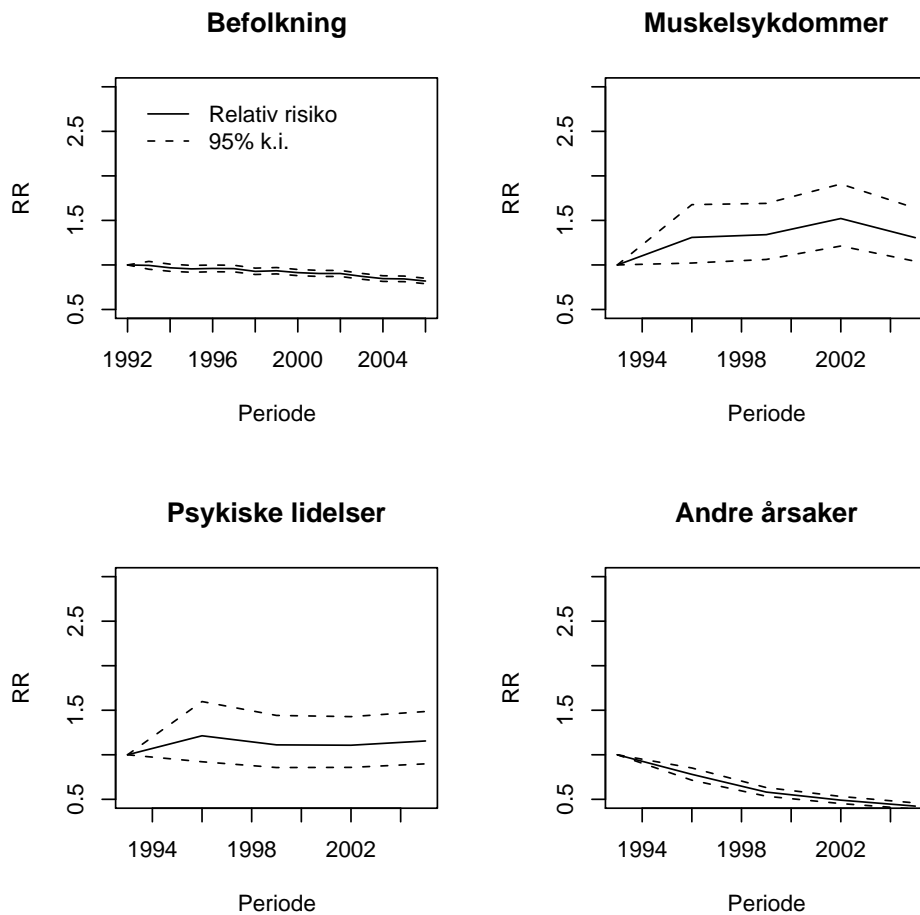
Tabell 5.3.4: *Deviansanalyser menn.*

5.4 Effekt av periode

Periode er den siste forklaringsvariabelen som tas i betraktning. Som tidligere skal vi med regresjonsmodellen $ADØD + PERIODE$ se hvordan denne forklaringsvariabelen påvirker dødeligheten. Dette gir dødsintensiteten $\mu_{a,p} = \exp(\beta_0 + \beta_a + \beta_p)$ for alder a ved død og død i periode p .

I figurene 5.4.1 og 5.4.2 vises den relative risikoen for død i periodene relativt til død i første periode. Dødsintensiteten for befolkningen viser seg å være synkende med perioden. Det samme er dødsintensiteten blant de uførepensjonerte i gruppa av andre årsaker. Reduksjonen er noe kraftigere for kvinner enn for menn. Den relative risikoen øker i første periode blant gruppa uførepensjonert grunnet psykiske lidelser. Deretter avtar den moderat. For de uføre med muskelsykdommer fortsetter dødsintensiteten å øke fram til siste periode. Begge uføregruppene har store konfidensintervaller for den relative risikoen, og vi kan derfor ikke påvise at endringene vi observerer er signifikante. De relative risikoene for gruppa med uførepensjonerte grunnet andre årsaker har små konfidensintervaller, og endres signifikant med periode.

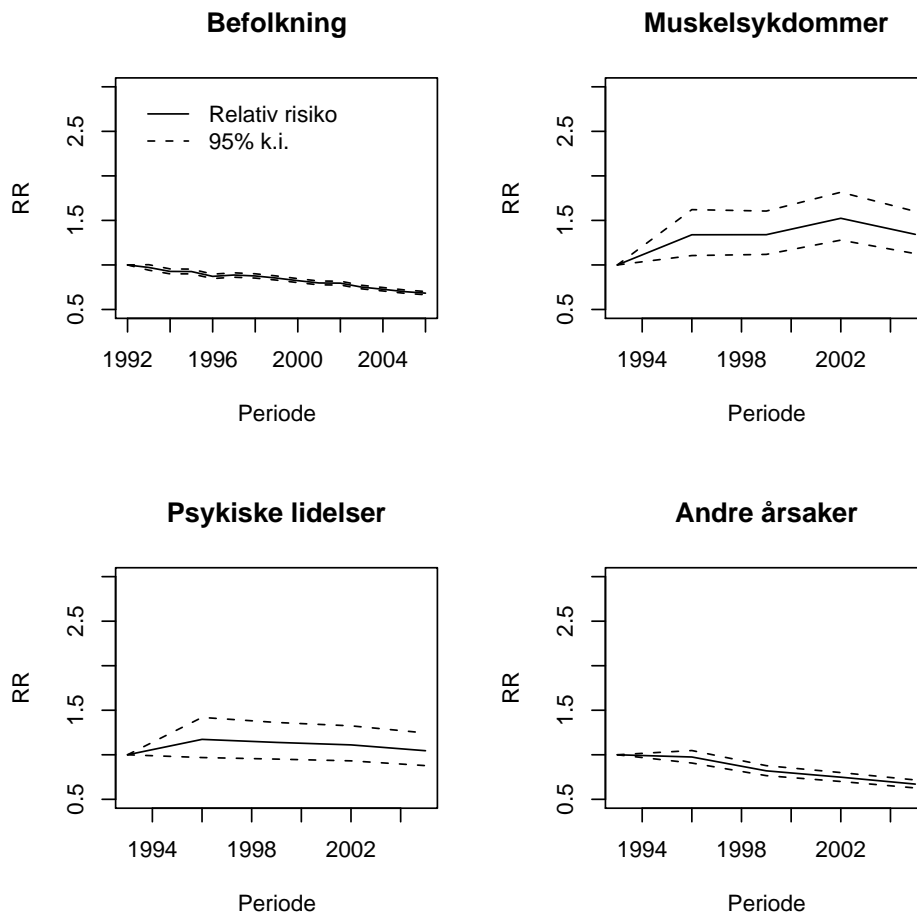
I figur 5.4.3 vises dødsintensitetene for 50-årige kvinner og menn over ulike perioder. Dødsintensiteten er betraktelig mye høyere for gruppa av uførepensjonerte grunnet andre årsaker enn for befolkningen. Kvinnene med psykiske lidelser har en lavere dødsintensitet enn hva mennene har. Vi ser også at gruppa av uførepensjonerte grunnet muskelsykdommer har, i de tidligste periodene, dødsintensiteter i underkant av befolkningens. Fordi dødsintensiteten er økende, overskrider den etterhvert befolkningens for begge kjønn.



Figur 5.4.1: Relativ risiko for død i ulike perioder relativt til død i året 1992 for befolkningen og perioden 1992-1994 for de uføre. Modell ADØD + PERIODE. Kvinner.

	B	P	M	A
Periode	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
92-94	1.00	1.00	1.00	1.00
95-97	0.96(0.92-1.00)	1.21(0.92-1.60)	1.31(1.02-1.68)	0.78(0.71-0.85)
98-00	0.93(0.90-0.97)	1.11(0.86-1.44)	1.34(1.06-1.69)	0.58(0.53-0.63)
01-03	0.90(0.87-0.94)	1.11(0.86-1.43)	1.52(1.21-1.91)	0.49(0.45-0.53)
04-06	0.84(0.81-0.88)	1.16(0.90-1.49)	1.31(1.04-1.64)	0.42(0.39-0.46)

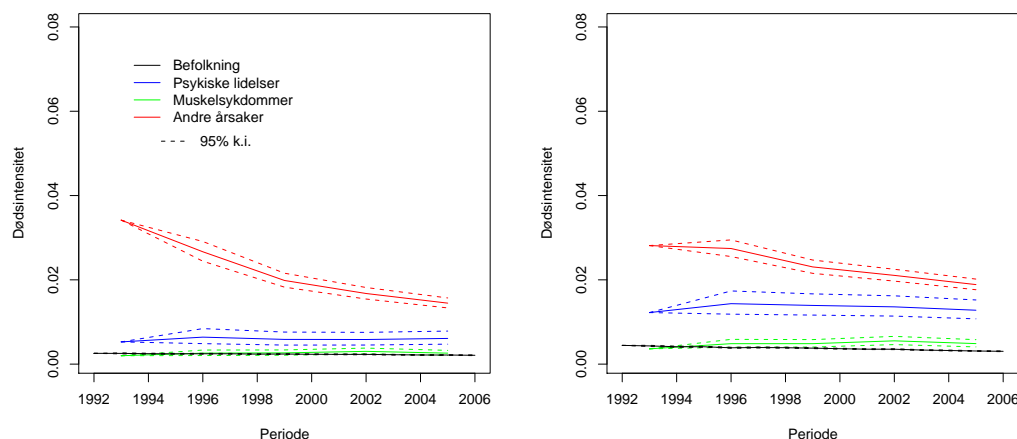
Tabell 5.4.1: Relativ risiko for død i ulike perioder relativt til død i året 1992 for befolkningen og perioden 1992-1994 for de uføre. Modell ADØD + PERIODE. Kvinner.



Figur 5.4.2: Relativ risiko for død i ulike perioder relativt til død i året 1992 for befolkningen og perioden 1992-1994 for de uføre. Modell ADØD + PERIODE. Menn.

	B	P	M	A
Periode	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
92-94	1.00	1.00	1.00	1.00
95-97	0.87(0.85-0.90)	1.17(0.97-1.42)	1.34(1.11-1.62)	0.98(0.91-1.05)
98-00	0.85(0.83-0.88)	1.14(0.95-1.36)	1.34(1.12-1.60)	0.82(0.77-0.88)
01-03	0.80(0.77-0.82)	1.11(0.93-1.32)	1.52(1.28-1.82)	0.75(0.70-0.80)
04-06	0.70(0.68-0.72)	1.05(0.88-1.24)	1.34(1.13-1.60)	0.67(0.63-0.72)

Tabell 5.4.2: Relativ risiko for død i ulike perioder relativt til død i året 1992 for befolkningen og perioden 1992-1994 for de uføre. Modell ADØD + PERIODE. Menn.



Figur 5.4.3: Dødsintensitet for en henholdsvis 50-årig kvinne og 50-årig mann, ved død i periodene mellom 1992 og 2006. Modell ADØD + PERIODE.

Deviansanalysene vises i tabellene 5.4.3 og 5.4.4. Vi ser at vi ved å utvide modellen ADØD med forklaringsvariabelen periode, reduserer deviansen signifikant for befolkningen, de uførepensjonerte grunnet muskelsykdommer og for de uførepensjonerte grunnet andre årsaker. Interaksjonen mellom alder ved død og perioden da dødsfallet skjer er også signifikant i disse gruppene. For de uføre grunnet psykiske lidelser er derimot periode ikke en signifikant forklaringsvariabel. Blant de uførepensjonerte kvinnene i denne gruppa kan vi heller ikke påvise noen interaksjon mellom de to forklaringsvariablene. Endringen i deviansen ved å tilpasse et interaksjonsledd er heller ikke stor for mennene, noe som gir en relativt stor p-verdi.

Årsak	Modell	Res.df	Res.dev	df	Endr.dev	p-verdi
P	1	321	981.88			
	+ADØD	310	475.80	11	506.08	<0.0001
	+PERIODE	306	472.90	4	2.89	0.58
	+ADØD*PERIODE	269	428.30	37	44.61	0.18
M	1	321	1894.12			
	+ADØD	310	599.26	11	1294.86	<0.0001
	+PERIODE	306	575.58	4	23.69	<0.0001
	+ADØD*PERIODE	269	477.24	37	98.34	<0.0001
A	1	321	3143.38			
	+ADØD	310	2351.79	11	791.59	<0.0001
	+PERIODE	306	1761.23	4	590.56	<0.0001
	+ADØD*PERIODE	269	1652.49	37	108.73	<0.0001

Tabell 5.4.3: Deviansanalyser kvinner.

Årsak	Modell	Res.df	Res.dev	df	Endr.dev	p-verdi
P	1	321	1402.15			
	+ADØD	310	455.17	11	946.98	<0.0001
	+PERIODE	306	447.53	4	7.63	0.11
	+ADØD*PERIODE	269	392.11	37	55.43	0.03
M	1	321	1957.70			
	+ADØD	310	615.84	11	1341.86	<0.0001
	+PERIODE	306	584.29	4	31.55	<0.0001
	+ADØD*PERIODE	269	508.27	37	76.01	0.0001
A	1	321	2950.38			
	+ADØD	310	1449.28	11	1501.10	<0.0001
	+PERIODE	306	1167.37	4	281.90	<0.0001
	+ADØD*PERIODE	269	1024.63	37	142.74	<0.0001

Tabell 5.4.4: *Deviansanalyser menn.*

5.5 Effekt av tilgangår og tid fra tilgang

Modellen utvides nå til ADØD + TID FRA TILGANG + TILGANG for samtidig å kunne ta hensyn til de to forklaringsvariablene tid fra tilgang og tilgangår. Dette gir dødsintensiteten $\mu_{a,t,f} = \exp(\beta_0 + \beta_a + \beta_t + \beta_f)$ for alder ved død a , tilgangår t og f år fra tilgang.

Først ser vi på hvordan effekten av tilgangår i modellen ADØD + TID FRA TILGANG endres når vi også inkluderer forklaringsvariabelen tilgangår. Deretter vurderes det motsatte tilfellet, nemlig hvordan effekten av tid fra tilgang endres i modellen ADØD + TILGANG ved å inkludere tid fra tilgang.

Dersom relativ risiko for død etter ulik tid fra tilgang relativt til død i tilgangår og for død i ulike tilgangår relativt til død i 92-93 fra modellen ADØD + TID FRA TILGANG + TILGANG, er signifikant forskjellig fra relativ risiko funnet fra modellene ADØD + TID FRA TILGANG og ADØD + TILGANG, kan vi si litt mer om en slik utvidelse av modellen er nødvendig.

Ved å inkludere tilgangår i modellen ADØD + TID FRA TILGANG, blir den relative risikoen for tid fra tilgang lavere. Dette gjelder for begge kjønn og for alle tre uføreårsaker. Forskjellen i de to modellene blir også større ettersom årene fra tilgang øker. Fra figurene 5.5.1 og 5.5.2 ser vi at forskjellene i den relative risikoen for død for gruppa med psykiske lidelser og de med muskelsykdommer etterhvert blir av en viss størrelse. Allikevel er de relative risikoene fra modellen med tre forklaringsvariabler inneholdt i konfidensintervallet til modellen med kun to forklaringsvariabler. For gruppa uførepensjonerte menn grunnet andre årsaker, er forskjellene mellom de to modellene som vises i figuren ikke spesielt stor. Derimot er forskjellene i de relative risikoene signifikant forskjellige, siden det i denne gruppa er spesielt små konfidensintervaller. Det samme ser vi for kvinner. Den relative risikoen listes i tabell 5.5.3 og 5.5.4. Disse sammenlignes med relativ risiko for modellen med kun to forklaringsvariabler i tabellene 5.3.1 og 5.3.2.

For modellen ADØD + TILGANG vil det unntaksvis ikke gi signifikante endringer i den

relative risikoen for tilgangssår ved også å inkludere forklaringsvariabelen tid fra tilgang i modellen. Dette vil gjelde for de uførepensjonerte grunnet psykiske lidelser og muskelsykdommer. Unntakene vil være for siste tilgang i 2004-2005, samt for tilgang i 1998-1999 for kvinner. For de uføre grunnet andre årsaker observerer vi så store forskjeller i relativ risiko at resultatene er signifikant forskjellige. Den relative risikoen listes i tabellene 5.5.1 og 5.5.2 og sammenlignes med tabellene 5.2.1 og 5.2.2.

Deviansen endrer seg signifikant ved å inkludere forklaringsvariabelen tilgang i modellen $ADØD + TID\ FRA\ TILGANG$ for de uførepensjonerte grunnet psykiske lidelser og muskelsykdommer. For gruppa av uførepensjonerte grunnet andre årsaker har vi også sterkt signifikante deviansendringer. At denne gruppa har en kraftig økning i dødsintensiteten for tilgang i årene 2004-2005, kan forklare hvorfor deviansen endres kraftig.

Signifikante interaksjoner finnes kun mellom tilgangssår og tid fra tilgang for gruppa av uførepensjonerte grunnet andre årsaker. At vi observerer slike interaksjoner kompliserer resultatene våre. Det er imidlertid vanskelig å studere disse interaksjonene nærmere på grunn av formen dataene foreligger på.

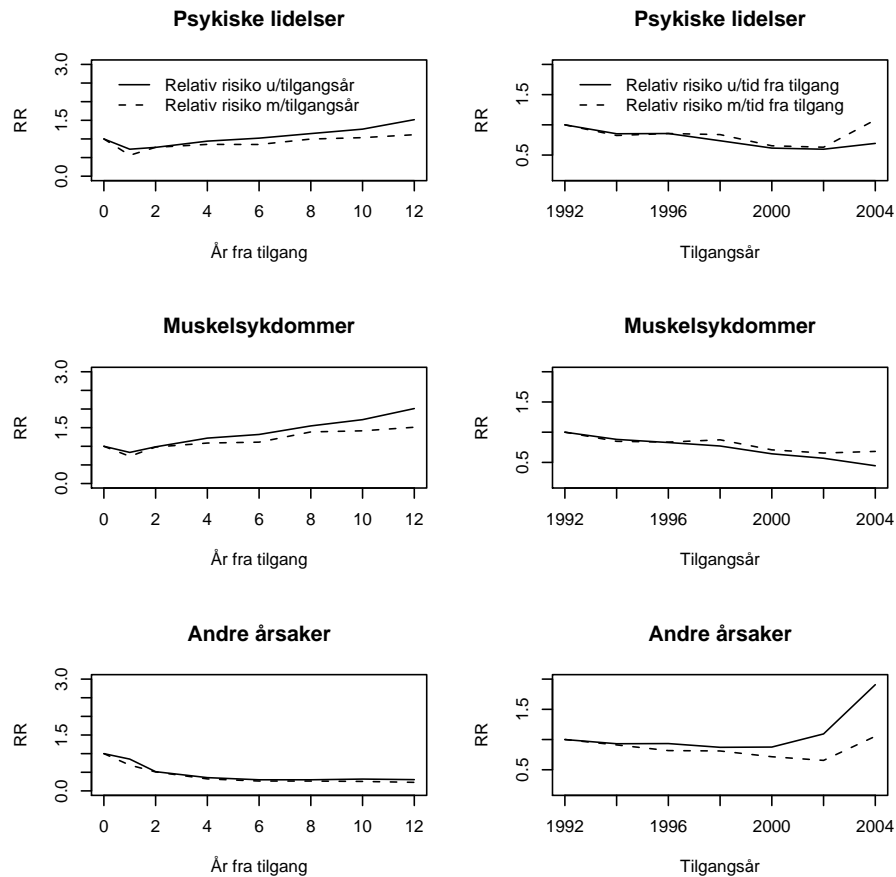
Det er verdt å merke seg at for de tre forklaringsvariablene tid fra tilgang, tilgangssår og periode, vil en av variablene kunne finnes ut ifra de to andre. Det gir derfor ikke mening å utvide modellen med nok en forklaringsvariabel. På grunn av formen dataene har, vil vi allerede ha innbyrdes avhengigheter i to av forklaringsvariablene, da vi for eksempel ikke kan ha tid siden tilgang på 10 år for personer med tilgang i 2004-2005.

	P	M	A
Tilgangssår	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
92-93	1.00	1.00	1.00
94-95	0.82 (0.71-0.96)	0.85 (0.75-0.96)	0.91 (0.84-0.99)
96-97	0.86 (0.72-1.02)	0.84 (0.73-0.96)	0.82 (0.75-0.89)
98-99	0.84 (0.72-0.98)	0.87 (0.77-0.99)	0.81 (0.75-0.87)
00-01	0.65 (0.54-0.79)	0.71 (0.60-0.83)	0.72 (0.65-0.79)
02-03	0.63 (0.49-0.80)	0.66 (0.53-0.81)	0.66 (0.59-0.73)
04-05	1.08 (0.77-1.52)	0.68 (0.49-0.95)	1.05 (0.95-1.17)

Tabell 5.5.1: *Relativ risiko for død i ulike tilgangssår relativt til død i 92-93. Kvinner med modellen $ADØD + TILGANGSÅR + TID\ FRA\ TILG$.*

	P	M	A
Tilgangssår	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
92-93	1.00	1.00	1.00
94-95	0.87 (0.78-0.98)	0.88 (0.80-0.97)	0.86 (0.81-0.91)
96-97	0.96 (0.85-1.09)	0.91 (0.81-1.02)	0.80 (0.75-0.85)
98-99	0.87 (0.78-0.98)	0.81 (0.73-0.91)	0.82 (0.78-0.87)
00-01	0.75 (0.66-0.86)	0.83 (0.72-0.94)	0.74 (0.69-0.79)
02-03	0.80 (0.68-0.95)	0.71 (0.59-0.85)	0.67 (0.61-0.73)
04-05	0.81 (0.64-1.02)	1.08 (0.85-1.37)	0.97 (0.88-1.06)

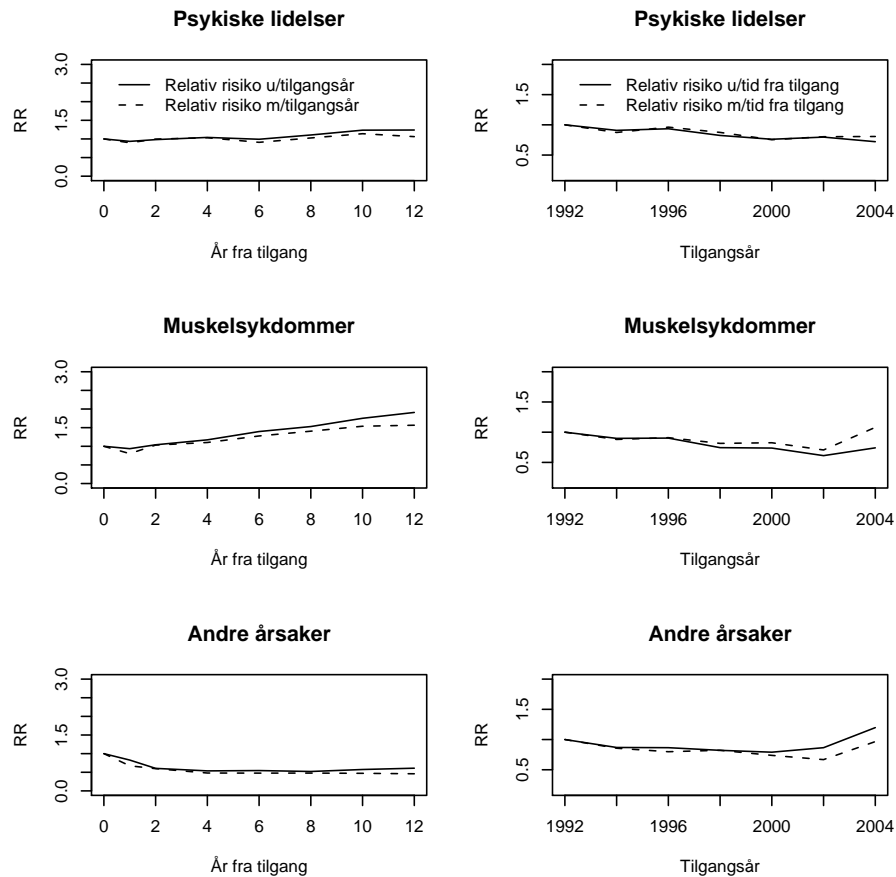
Tabell 5.5.2: *Relativ risiko for død i ulike tilgangssår relativt til død i 92-93. Menn med modellen $ADØD + TILGANGSÅR + TID\ FRA\ TILG$.*



Figur 5.5.1: Venstre kolonne: Relativ risiko for død etter ulik tid fra tilgang i modellene ADØD + TID FRA TILG og ADØD + TILGANG + TID FRA TILG. Høyre kolonne: Relativ risiko for død for ulike tilgangår i ADØD + TILGANG og ADØD + TILGANG + TID FRA TILG. Kvinner.

	P	M	A
Tid fra tilgang	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
0	1.00	1.00	1.00
1	0.56 (0.42-0.75)	0.73 (0.56-0.95)	0.69 (0.62-0.77)
2	0.78 (0.63-0.95)	0.98 (0.80-1.20)	0.51 (0.47-0.55)
4	0.85 (0.68-1.07)	1.09 (0.88-1.35)	0.32 (0.29-0.35)
6	0.85 (0.67-1.08)	1.11 (0.89-1.39)	0.26 (0.24-0.29)
8	1.00 (0.79-1.25)	1.39 (1.11-1.72)	0.26 (0.24-0.29)
10	1.03 (0.81-1.32)	1.42 (1.12-1.78)	0.25 (0.23-0.28)
12	1.11 (0.83-1.50)	1.51 (1.16-1.97)	0.23 (0.19-0.27)

Tabell 5.5.3: Relativ risiko for død etter ulik tid fra tilgang relativt til død i tilgangår. Modellen ADØD + TILGANGSÅR + TID FRA TILG. Kvinner.



Figur 5.5.2: Venstre kolonne: Relativ risiko for død etter ulik tid fra tilgang i modellene $ADØD + TID\ FRA\ TILG$ og $ADØD + TILGANG + TID\ FRA\ TILG$. Høyre kolonne: Relativ risiko for død i ulike tilgangår i $ADØD + TILGANG$ og $ADØD + TILGANG + TID\ FRA\ TILG$. Menn.

	P	M	A
Tid fra tilgang	RR(95% k.i.)	RR(95% k.i.)	RR(95% k.i.)
0	1.00	1.00	1.00
1	0.90 (0.73-1.10)	0.80 (0.64-1.01)	0.67 (0.62-0.74)
2	1.00 (0.86-1.16)	1.03 (0.87-1.23)	0.60 (0.56-0.64)
4	1.03 (0.87-1.22)	1.10 (0.91-1.33)	0.48 (0.45-0.52)
6	0.91 (0.76-1.08)	1.28 (1.06-1.55)	0.48 (0.44-0.52)
8	1.03 (0.86-1.22)	1.41 (1.16-1.70)	0.48 (0.44-0.52)
10	1.14 (0.95-1.38)	1.54 (1.26-1.88)	0.47 (0.43-0.51)
12	1.06 (0.84-1.34)	1.57 (1.24-1.97)	0.46 (0.41-0.52)

Tabell 5.5.4: Relativ risiko for død etter ulik tid fra tilgang relativt til død i tilgangår. Modellen $ADØD + TILGANGSÅR + TID\ FRA\ TILG$. Menn.

Årsak	Modell	Res.df	Res.dev	df	Endr.dev	p-verdi
P	1	321	981.88			
	+ADØD	310	475.80	11	506.08	<0.0001
	+TID FRA TILG	303	403.28	7	72.52	<0.0001
	+TILGANG	297	377.70	6	25.58	0.0003
	+TID FRA TILG* TILGANG	288	365.82	9	11.88	0.22
M	1	321	1894.12			
	+ADØD	310	599.26	11	1294.86	<0.0001
	+TID FRA TILG	303	463.78	7	135.48	<0.0001
	+TILGANG	297	436.61	6	27.17	0.0001
	+TID FRA TILG* TILGANG	288	421.50	9	15.11	0.09
A	1	321	3143.38			
	+ADØD	310	2351.79	11	791.59	<0.0001
	+TID FRA TILG	303	810.83	7	1540.96	<0.0001
	+TILGANG	297	699.84	6	110.98	<0.0001
	+TID FRA TILG* TILGANG	288	608.98	9	90.86	<0.0001

Tabell 5.5.5: *Deviansanalyser kvinner.*

Årsak	Modell	Res.df	Res.dev	df	Endr.dev	p-verdi
P	1	321	1402.15			
	+ADØD	310	455.17	11	946.98	<0.0001
	+TID FRA TILG	303	430.94	7	24.23	0.001
	+TILGANG	297	406.21	6	24.73	0.0004
	+TID FRA TILG * TILGANG	288	396.26	9	9.95	0.35
M	1	321	1957.70			
	+ADØD	310	615.84	11	1341.86	<0.0001
	+TID FRA TILG	303	473.34	7	142.50	<0.0001
	+TILGANG	297	445.53	6	27.81	0.0001
	+TID FRA TILG * TILGANG	288	428.69	9	16.84	0.05
A	1	321	2950.38			
	+ADØD	310	1449.28	11	1501.10	<0.0001
	+TID FRA TILG	303	886.97	7	562.30	<0.0001
	+TILGANG	297	760.98	6	126.00	<0.0001
	+TID FRA TILG * TILGANG	288	649.09	9	111.88	<0.0001

Tabell 5.5.6: *Deviansanalyser menn.*

5.6 Sammenligning av effekter

Fram til nå har vi presentert en rekke modeller med ulike forklaringsvariabler for å beskrive uføredødeligheten. Ved å benytte Akaike's information criterion(AIC) fra (3.21), kan vi sammenligne modeller med et ulikt antall forklaringsvariabler. I tabell 5.6.1 ser vi at AIC reduseres kraftig for alle uføreårsaker og begge kjønn dersom vi benytter forklaringsvariablen alder ved død i modellen ADØD.

For samtlige av modellene reduseres AIC ved å tilpasse en ekstra forklaringsvariabel. Det er forklaringsvariablen tid fra tilgang som bidrar til størst reduksjon i AIC for alle tilfellene, utenom for menn med psykiske lidelser, hvor tilgangssår gir størst reduksjon. Når forklaringsvariablen periode inkluderes i modellen ADØD, bidrar denne til en svak reduksjon i AIC. Dette gjelder for alle uføreårsakene, unntatt de uførepensjonerte grunnet psykiske lidelser, hvor tilpasningen av denne forklaringsvariablen ikke gir noen gevinst i form av AIC-reduksjon.

Dersom vi utvider modellen til å inneholde tre forklaringsvariabler med modellen ADØD + TID FRA TILGANG + TILGANG, ser vi at AIC igjen reduseres noe for alle utenom de med psykiske lidelser. For de uførepensjonerte kvinnene grunnet andre årsaker ser vi at reduksjonen i AIC er kraftigere når vi inkluderer tid fra tilgang enn når alder ved død inkluderes. Altså er tid fra tilgang en viktigere variabel enn alder ved død i forklaringen av dødsintensiteten til denne gruppa av uføre.

Modellen ADØD + TID FRA TILGANG + TILGANG kan vi da anta at er tilstrekkelig for å beskrive dødsintensiteten for de uføre grunnet muskelsykdommer og psykiske lidelser, da denne gir lavest AIC. For de uførepensjonerte grunnet psykiske lidelser er muligens ADØD + TILGANG tilstrekkelig. For de uførepensjonerte grunnet andre årsaker, vil modellen hvor vi interaksjonen tilpasses, altså ADØD + TID FRA TILGANG + TILGANG, best beskrive dødsintensiteten.

Kjønn	Modell	df	AIC		
			Psykiske lidelser	Muskel-sykdommer	Andre årsaker
K	1	1	2094	3046	4754
	ADØD	12	1610	1773	3984
	ADØD + TILG	18	1566	1694	3742
	ADØD + TID FRA TILGANG	19	1551	1652	2457
	ADØD + PERIODE	16	1615	1758	3401
	ADØD + TID FRA TILG + TILGANG	25	1538	1637	2358
	ADØD + TID FRA TILG * TILGANG	34	1544	1639	2285
M	1	1	2729	3142	4715
	ADØD	12	1804	1822	3236
	ADØD + TILGANG	18	1780	1758	3079
	ADØD + TID FRA TILG	19	1793	1694	2687
	ADØD + PERIODE	16	1804	1799	2962
	ADØD + TID FRA TILG + TILGANG	25	1781	1678	2573
	ADØD + TID FRA TILG * TILGANG	34	1789	1679	2479

Tabell 5.6.1: AIC for kvinner og menn.

Kapittel 6

Sensitivitet

I modelleringen av dødeligheten finnes det mange kilder til feilestimering. I dette kapittelet skal vi vurdere noen av de mest opplagte årsakene til feil. Dette vil først og fremst omfatte behandlingen av rådataene, samt usikkerhet i tilpasningene og antagelsene som foretas. Vi vil også forsøke å kvantifisere noen av usikkerhetsmomentene.

6.1 Metodikk for befolkningsdataene

I oppgaven er det antatt 15 års oppfølging på befolkningsdataene. Dette er, som tidligere nevnt, riktig kun når vi har en lukket befolkning. Det vil altså si at emigrasjon og immigrasjon ikke forekommer, og at all reduksjon i befolkningsstørrelsen derfor skjer som følge av dødsfall. Allikevel har det blitt foretatt oppfølging, slik at befolkningsdataene lettere kan sammenlignes med uføredataene. Vi skal nå se i hvor stor grad denne antagelsen influerer resultatene. Den anvendte metoden sammenlignes med en metode der antallet døde divideres med den faktiske befolkningsstørrelsen i det inneværende året uten noen form for oppfølging.

I tabellen 6.1.1 vises de optimale verdiene tilpasset alder 18 til 81 år for metoden med oppfølging, og alder 18 til 67 for de hvor vi ikke foretar noen oppfølging.

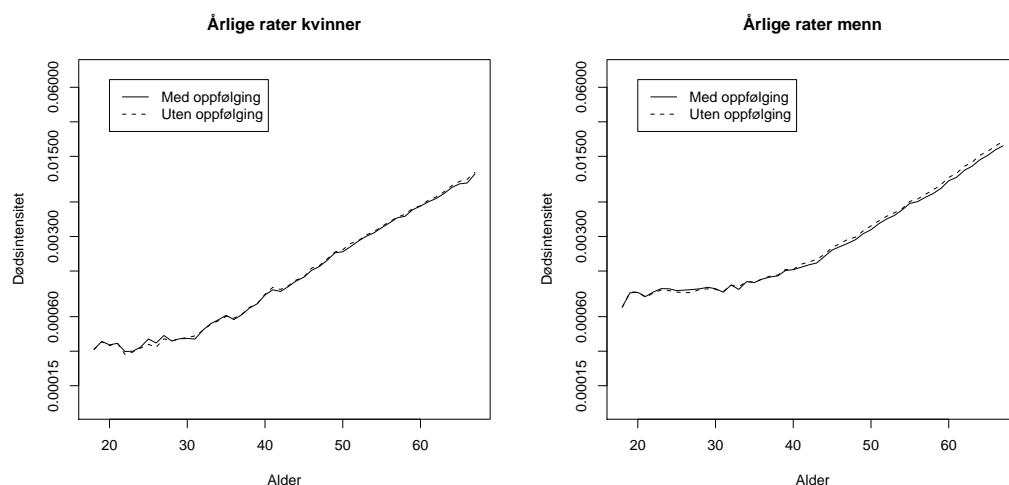
	Med oppfølging		Uten oppfølging	
	$\hat{\beta}$	\hat{c}	$\hat{\beta}$	\hat{c}
Kvinner	2.57E-5	1.0937	2.64E-5	1.0906
Menn	5.29E-5	1.0907	5.52E-5	1.0825

Tabell 6.1.1: *Optimale verdier for Gompertz-tilpasningen med og uten oppfølging.*

Den årlige veksten, \hat{c} , er for begge kjønn lavere i metoden uten oppfølging. Derimot er $\hat{\beta}$ -verdien noe høyere. Forskjellene i tilpasningen av Gompertz-parameterne er imidlertid veldig små.

De årlige ratene fra de to metodene vises i figur 6.1.1, og ser ikke ut til å gi spesielt forskjellige resultater. Imidlertid kan vi se små tendenser til at vi med oppfølgingen overestimerer dødsintensiteten for befolkningen i alder opp til 30 år. I samme grad ser vi ut til å underestimere befolkningsdødeligheten når befolkningen er i alder fra 40 år og oppover.

Relativ risiko for dødsintensiteten blant uføre relativt til befolkningen ble vist i figur 5.1.2, og listet i tabellene 5.1.1 og 5.1.2. Disse vil da altså i ung alder være noe høyere enn for metoden med oppfølging, og tilsvarende vil de relative risikoene for død være mindre i høyere alder. Feilestimeringen som gjøres vil være av beskjeden størrelse.



Figur 6.1.1: Sammenlikning av årlige rater for befolkningsdata med oppfølging og befolkningsdata uten oppfølging, for henholdsvis kvinner og menn i alderen 18 til 67 år. Logaritmisk skala.

6.2 Aggregering av data

Fordi det forekommer et lite antall dødsfall hvert år for de ulike uføregruppene, er disse dataene som kjent aggregert. Dette innebærer eksempelvis at effekter fra enkelte årskull kan bli utlignet av bidragene fra andre årskull som inngår i den oppaggregerte gruppa.

Vi skal nå se på forskjellene i dødsintensiteten fra befolkningsdata på årlig skala og til befolkningsdata aggregert tilsvarende uføredataene. Aldersgruppene settes sammen til grupper av ulike størrelser, mens to og to av det vi har benevnt tilgangså, slås sammen. I tillegg grupperes dødsårene i grupper på tre år. På grunn av denne aggregeringen, vil de som dør i et av dødsårsintervallene og de som overlever, få gjennomsnittlige observasjonstider i de ulike intervallene.

De anslåtte observasjonstidene beregnes ut i fra simuleringer, slik det blir beskrevet i tillegg A. Disse vises i tabell 6.2.2 og er noe forskjellig fra observasjonstidene beregnet for uføredataene. Dette er fordi befolkningsstørrelsen tas fra begynnelsen av hvert år, mens de uføre har tilgang til uførhet gjennom hele intervallet.

De optimale tilpassede verdiene for de aggregerte dataene listes i tabell 6.2.1. Det ser ikke ut til at parameterne tilpasset de aggregerte dataene er svært forskjellige fra de estimert fra årlige data. Det ser ut til at vekstraten er noe høyere, mens $\hat{\beta}$ -verdien er noe lavere. Dette gjelder for begge kjønn.

De årlige ratene fra de oppaggregerte dataene sammenlignes med Gompertz-tilpasningen i figur 6.2.1. For at svingningene i de årlige ratene ikke skal bli for store, glatter vi ratene ved å se på 5-årige aldersspenn av alder ved død. Dette kan sammenlignes med de fullstendige dataene i figur 4.1.2, og tilpasningene ser tilsvarende ut.

De årlige ratene for de aggregerte dataene er tilnærmet log-lineære gjennom hele aldersintervallet vi observerer for kvinner. Dette gjør at dødsintensiteten fra den tilpassede Gompertz-fordelingen gir en grei tilpasning til de årlige ratene. Tilpasningen er ikke like god for menn.

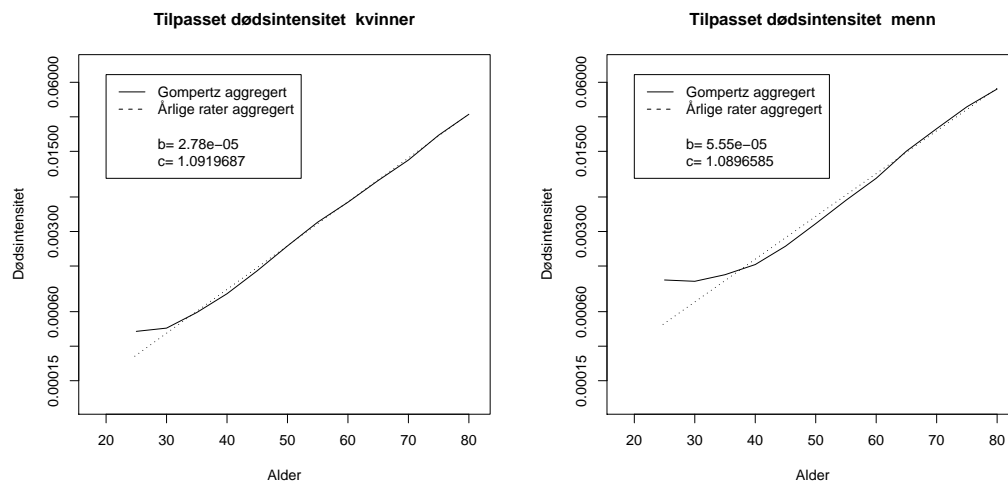
Vi sammenligner også de to dataformatene ved å benytte Poisson-regresjon. De årlige ratene med 5-års intervaller på alder ved død, tilsvarer regresjonsanalysen med modellen ADØD, slik vi beskrev i eksempel 3.3.5. Det ser ikke ut til at dødsintensitetene fra de aggregerte dataene som vises i figur 6.2.2 avviker nevneverdig mye fra årsdataene. Imidlertid kan det synes at de aggregerte dataene overestimerer dødsintensiteten noe i ung alder, og svakt underestimerer den fra 60 år og oppover. Spesielt ser vi at aggregeringen til den store aldersgruppa fra 18 til 35 er kilde til de største avvikene.

	Aggregerte data		Årsdata	
	$\hat{\beta}$	\hat{c}	$\hat{\beta}$	\hat{c}
Kvinner	2.78E-05	1.0920	2.57E-05	1.0937
Menn	5.55E-05	1.0897	5.29E-05	1.0907

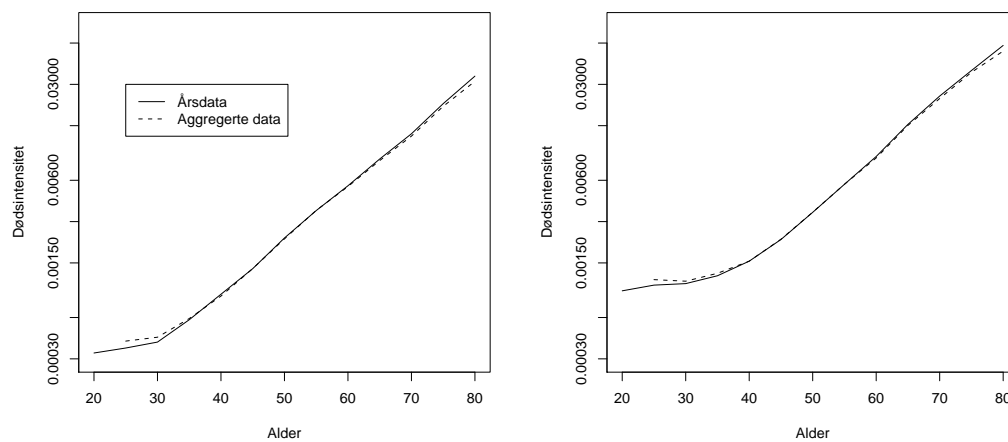
Tabell 6.2.1: *Optimale verdier for Gompertz-tilpasningen for aggregerte data og årsdata.*

Tilgangså	Dødsår				
	1992 – 1994	1995 – 1997	1998 – 2000	2001 – 2003	2004 – 2006
1992 – 1993	1.25/2.50	1.50/3.00	1.50/3.00	1.50/3.00	1.50/3.00
1994 – 1995	0.25/0.50	1.50/3.00	1.50/3.00	1.50/3.00	1.50/3.00
1996 – 1997		0.75/1.50	1.50/3.00	1.50/3.00	1.50/3.00
1998 – 1999			1.25/2.50	1.50/3.00	1.50/3.00
2000 – 2001			0.25/0.50	1.50/3.00	1.50/3.00
2002 – 2003				0.75/1.50	1.50/3.00
2004 – 2005					1.25/2.50

Tabell 6.2.2: *Anslåtte observasjonstider i kohortene for aggregert befolkning. Gitt død/gitt overlevelse.*



Figur 6.2.1: Dødsintensitet for kvinner og menn. Data på aggregert form sammenlignes med Gompertz-resultatene for årsdata. Logaritmisk skala.

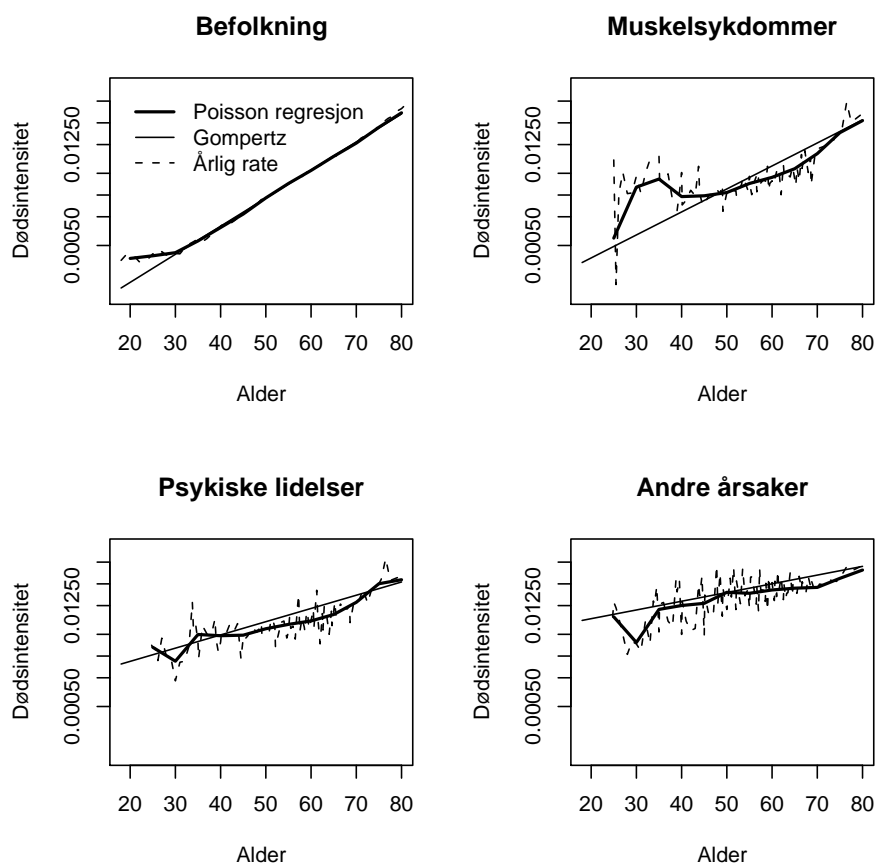


Figur 6.2.2: Dødsintensitet for befolkningen gitt ved årsdata og aggregert tilsvarende uføredataene for henholdsvis kvinner og menn. Logaritmisk skala. Modell ADØD.

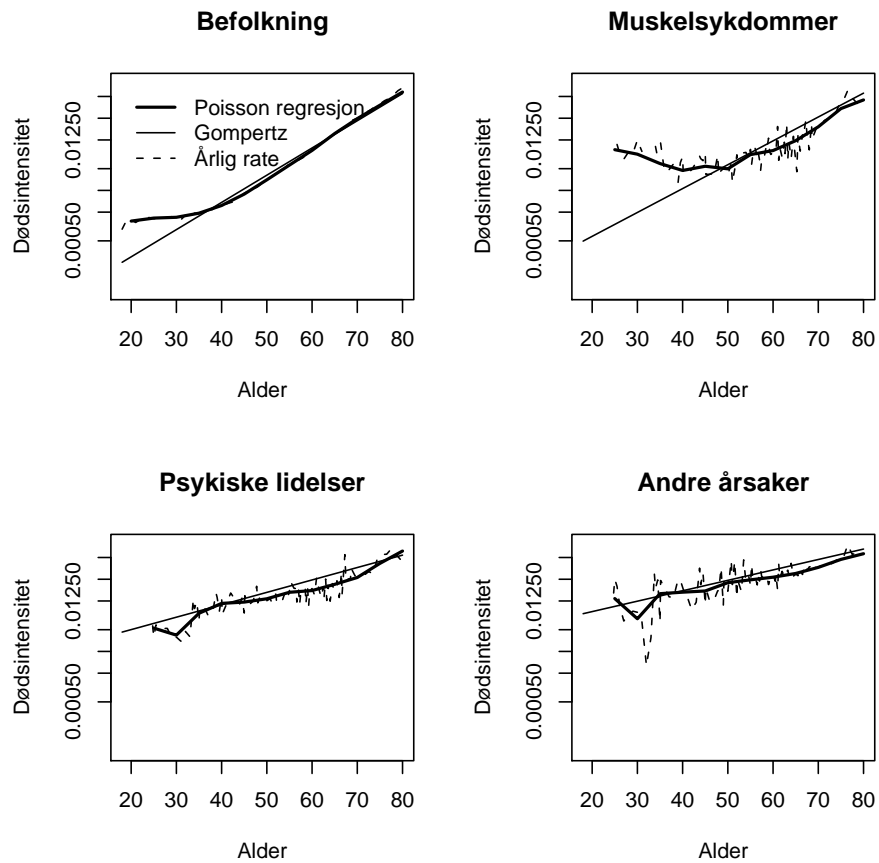
6.3 Sammenligning av metoder

For å estimere dødsintensiteten har to ulike metoder blitt benyttet. Dette er Poisson-regresjon og optimering av en Gompertz-fordeling. Gompertz-fordelingen gir log-lineære dødsintensiteter. Som vi så av resultatene ga dette tildels store feil i tilpasningen av dødsintensiteten i ung alder. Dødsintensiteten beregnet ut ifra parameterne fra optimeringen av Gompertz-fordelingen i kapittel 4, vises sammen med dødsintensitetene fra regresjonsmodellen ADØD og de årlige ratene i figurene 6.3.1 og 6.3.2.

Fra figurene ser vi at Poisson-regresjonen tilpasser dødsintensiteten i ung alder bedre enn hva Gompertz-fordelingen gjør. Etter fylte 40 år ser ikke de to tilpassede dødsintensitetene ut til å være urovekkende forskjellige. Imidlertid ser vi tendenser til at dødsintensiteten fra Gompertz-fordelingen, i endel av årsaksgruppene, overestimerer dødsintensiteten når individene er mellom 40 og 70 år. Tilsvarende kan det synes at dødsintensiteten underestimeres når individene er over 70 år.



Figur 6.3.1: Sammenligning av dødsintensitet i alle oppfølgingsår ved optimering av Gompertz og fra Poisson-regresjon. Befolkningsdata, kvinner. Logaritmisk skala. Modell ADØD.



Figur 6.3.2: Sammenligning av dødsintensitet i alle oppfølgingsår ved optimering av Gompertz og fra Poisson-regresjon. Befolkningsdata, menn. Logaritmisk skala. Modell ADØD.

6.4 Imputasjoner

Som beskrevet i kapittel 2.4, blir tilfeller hvor antall døde uføre ligger mellom 1 og 4, intervallsensurerte. For alle data tilgjengelig fins det for hvert kjønn og uføreårsak totalt 322 aldersgrupperinger eller celler av døde uføre. Tabell 6.4.1 viser hvordan de sensurerte cellene fordeler seg mellom de ulike uføreårsakene.

Årsak	Kvinner	Menn
Psykiske lidelser	98	42
Muskelsykdommer	93	75
Andre årsaker	12	5
Totalt	203	122

Tabell 6.4.1: *Sensurerte celler.*

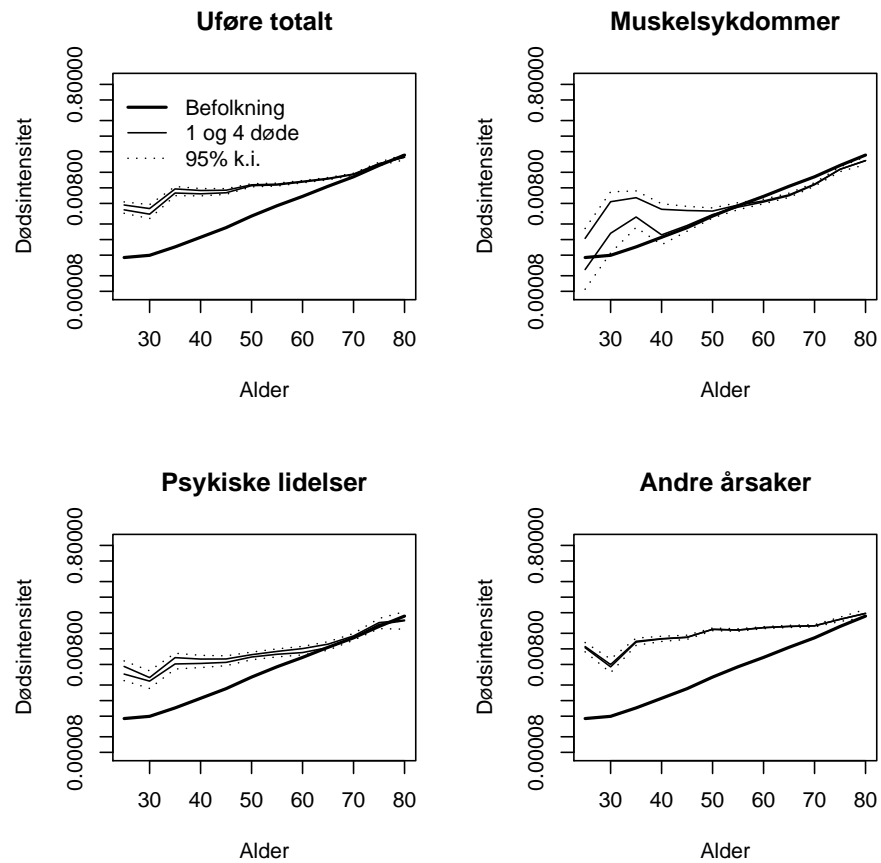
Dette betyr at omkring 21% av cellene for kvinner er sensurerte, mens det for menn er kun 13%. Blant menn er det flest sensurerte celler i dataene for uførepenjonerte grunnet muskelsykdommer. For kvinner er det mange sensurerte celler både i gruppa av uførepenjonerte grunnet muskelsykdommer og psykiske lidelser. Hovedvekten av sensureringen ligger for begge kjønn blant de unge uføre, hvor det naturlig nok er få som dør.

I beregningene som foretas i oppgaven, settes de sensurerte cellene til å ha en fast verdi på 3 døde. For å kunne foreta en slik antagelse, må vi vite mer om i hvor stor grad dødsintensiteten påvirkes av endringer i imputasjonsverdien. Ulik metodikk for valg av imputasjonsverdier diskuteres i inneværende seksjon.

6.4.1 Faste imputasjoner

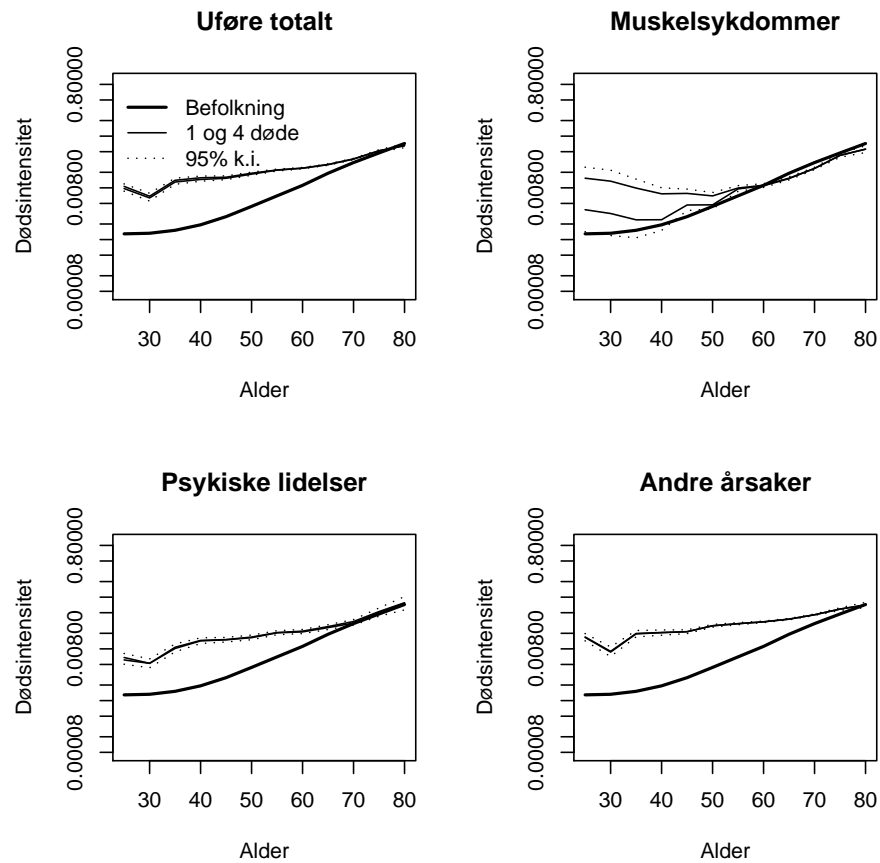
Først skal vi se hvordan dødsintensiteten påvirkes ved å benytte ytterpunktene i sensurintervallet, nemlig 1 og 4 døde, som faste imputasjonsverdier. Det vil altså si at vi vurderer den maksimale spredningen beregningene kan få. Det er ikke plausibelt å anta at spredningen er av en slik størrelse. En slik metode kan allikevel gi oss et foreløpig inntrykk av i hvilken grad spredningen gir utslag.

I figur 6.4.1 og 6.4.2 vises dødsintensiteten for de ulike uføregruppene estimert med regresjonsmodellen ADØD. Vi sammenligner med befolkningsdødeligheten. Spesielt ser vi at endringer i imputasjonsverdien fra 1 til 4 er utslagsgivende i ung alder for de muskelsyke av begge kjønn. Endringen i dødsintensiteten for kvinner med psykiske lidelser, er også merkbar. Derimot har endringen liten innvirkning på de uførepenjonerte grunnet andre årsaker. Det er spesielt i ung alder, hvor vi har få personer som dør, at endringer imputeringsverdien gir store forskjeller i dødsintensiteten. De relative endringene i dødsintensiteten fra imputasjonsverdier på 3 vises i tabellene 6.4.2 og 6.4.3.

Figur 6.4.1: *Maksimal spredning på dødsintensiteten. Modell ADØD. Kvinner.*

Alder	Uføre totalt		Psykiske lidelser		Muskelsykdommer		Andre årsaker	
	1 død	4 døde	1 død	4 døde	1 død	4 døde	1 død	4 døde
17.5-22.5	0.86	1.07	0.79	1.11	0.33	1.33	0.97	1.02
22.5-27.5	0.84	1.08	0.90	1.05	0.33	1.34	0.93	1.03
27.5-32.5	0.88	1.06	0.82	1.09	0.53	1.24	0.98	1.01
32.5-37.5	0.90	1.05	0.87	1.06	0.41	1.29	0.99	1.00
37.5-42.5	0.93	1.04	0.90	1.05	0.61	1.19	0.99	1.00
42.5-47.5	0.97	1.01	0.93	1.03	0.90	1.05	1.00	1.00
52.5-57.5	0.98	1.01	0.92	1.04	0.94	1.03	1.00	1.00
57.5-62.5	0.98	1.01	0.89	1.06	0.97	1.02	1.00	1.00
62.5-67.5	0.98	1.01	0.90	1.05	0.98	1.01	1.00	1.00
67.5-72.5	0.99	1.01	0.95	1.03	0.98	1.01	1.00	1.00
72.5-77.5	0.99	1.00	0.95	1.03	1.00	1.00	1.00	1.00
77.5-82.5	1.00	1.00	0.98	1.01	1.00	1.00	1.00	1.00

Tabell 6.4.2: *Endring i dødsintensiteten relativt til at de sensurerte cellene er antatt å være 3. Kvinner.*



Figur 6.4.2: *Maksimal spredning på dødsintensiteten. Modell ADØD. Menn.*

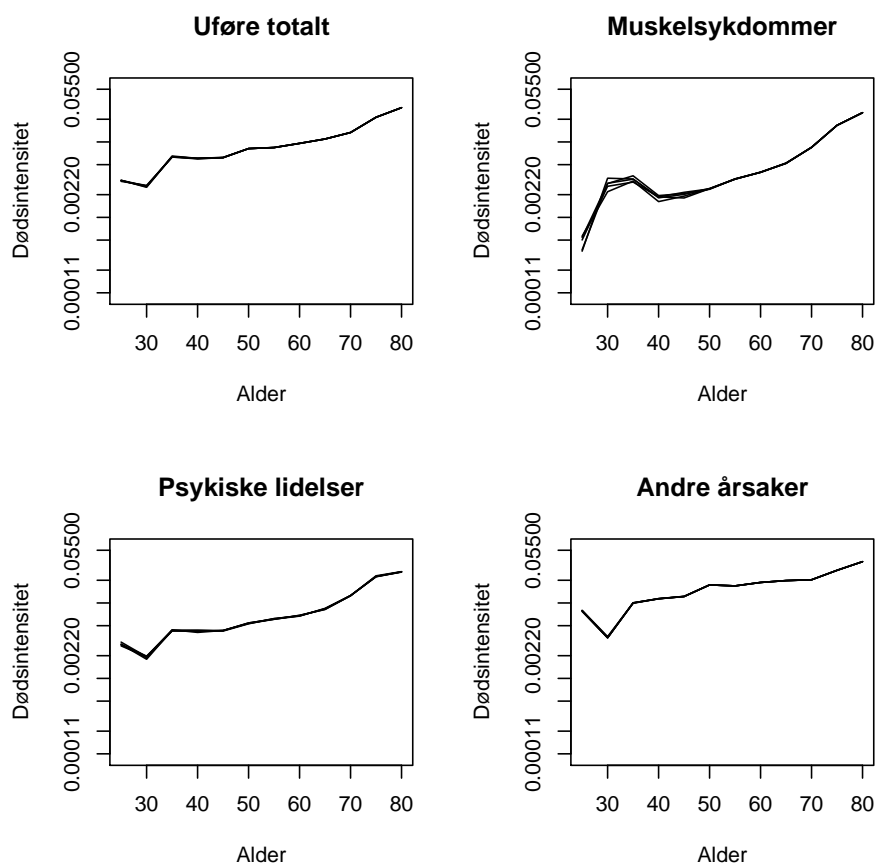
Alder	Uføre totalt		Psyriske lidelser		Muskelsykdommer		Andre årsaker	
	1 død	4 døde	1 død	4 døde	1 død	4 døde	1 død	4 døde
17.5-22.5	0.86	1.07	0.79	1.11	0.33	1.33	0.97	1.02
22.5-27.5	0.84	1.08	0.90	1.05	0.33	1.34	0.93	1.03
27.5-32.5	0.88	1.06	0.82	1.09	0.53	1.24	0.98	1.01
32.5-37.5	0.90	1.05	0.87	1.06	0.41	1.29	0.99	1.00
37.5-42.5	0.93	1.04	0.90	1.05	0.61	1.19	0.99	1.00
42.5-47.5	0.97	1.01	0.93	1.03	0.90	1.05	1.00	1.00
52.5-57.5	0.98	1.01	0.92	1.04	0.94	1.03	1.00	1.00
57.5-62.5	0.98	1.01	0.89	1.06	0.97	1.02	1.00	1.00
62.5-67.5	0.98	1.01	0.90	1.05	0.98	1.01	1.00	1.00
67.5-72.5	0.99	1.01	0.95	1.03	0.98	1.01	1.00	1.00
72.5-77.5	0.99	1.00	0.95	1.03	1.00	1.00	1.00	1.00
77.5-82.5	1.00	1.00	0.98	1.01	1.00	1.00	1.00	1.00

Tabell 6.4.3: *Endring i dødsintensiteten relativt til at de sensurerte cellene er antatt å være 3. Menn.*

6.4.2 Uniform trekking

En alternativ metode å behandle imputeringsproblematikken på, vil være å trekke hver av de imputerte verdiene tilfeldig fra en uniform fordeling, $\sim U(1, 4)$. Dødsintensitetene finnes ved å gjøre regresjonsanalyser på dataene med de nye imputasjonsverdiene. Ved å gjøre dette gjentatte ganger, kan vi få et inntrykk av hvor sensitive beregningene er for endringer i imputasjonsverdiene. Dette har blitt gjort for henholdsvis kvinner og menn i figur 6.4.3 og 6.4.4.

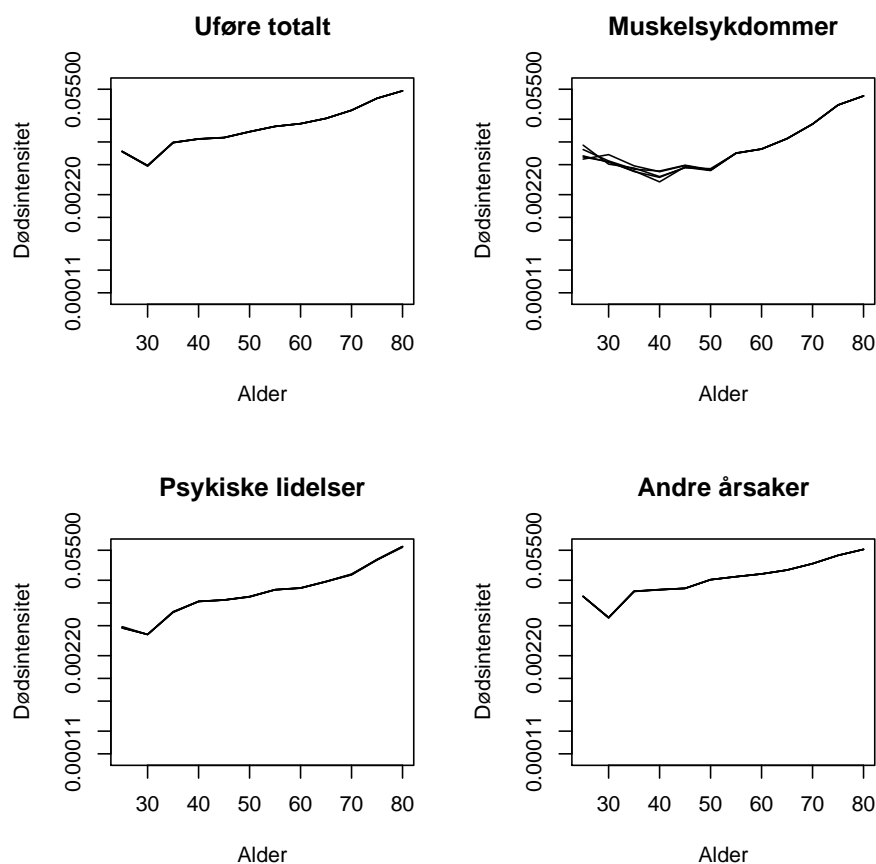
Vi ser også her at imputeringen først og fremst påvirker dødsintensiteten til de uføre-pensjonerte med muskelsykdommer. Å trekke tilfeldig imputeringsverdier, gir mindre avvik mellom dødsintensitetene som estimeres. Spesielt reduseres usikkerheten for de muskelsyke over 40 år.



Figur 6.4.3: Fire realisasjoner av dødsintensiteten hvor de imputerte verdiene trekkes $\sim U(1,4)$. Modell ADØD. Kvinner.

Alder	Uføre totalt	Psykiske lidelser	Muskelsykdommer	Andre årsaker
17.5-22.5	(0.97-1.03)	(0.93-1.07)	(0.74-1.26)	(0.98-1.02)
22.5-27.5	(0.95-1.05)	(0.95-1.05)	(0.57-1.43)	(0.96-1.04)
27.5-32.5	(0.96-1.04)	(0.95-1.05)	(0.78-1.22)	(0.99-1.01)
32.5-37.5	(0.98-1.02)	(0.97-1.03)	(0.78-1.22)	(0.99-1.01)
37.5-42.5	(0.99-1.01)	(0.98-1.02)	(0.89-1.11)	(1.00-1.00)
42.5-47.5	(0.99-1.01)	(0.98-1.02)	(0.96-1.04)	(1.00-1.00)
52.5-57.5	(1.00-1.00)	(0.98-1.02)	(0.99-1.01)	(1.00-1.00)
57.5-62.5	(1.00-1.00)	(0.98-1.02)	(1.00-1.00)	(1.00-1.00)
62.5-67.5	(1.00-1.00)	(0.98-1.02)	(0.99-1.01)	(1.00-1.00)
67.5-72.5	(1.00-1.00)	(0.98-1.02)	(1.00-1.00)	(1.00-1.00)
72.5-77.5	(1.00-1.00)	(0.97-1.03)	(1.00-1.00)	(1.00-1.00)
77.5-82.5	(1.00-1.00)	(0.99-1.01)	(1.00-1.00)	(1.00-1.00)

Tabell 6.4.4: Et 95% konfidensintervall på relativ endring i dødsintensiteten fra gjennomsnittlig dødsintensitet. Kvinner.



Figur 6.4.4: Fire realisasjoner av dødsintensiteten hvor de imputerte verdiene trekkes $\sim U(1,4)$. Modell ADØD. Menn.

Alder	Uføre totalt	Psykiske lidelser	Muskelsykdommer	Andre lidelser
17.5-22.5	(0.98-1.02)	(0.96-1.04)	(0.78-1.22)	(0.99-1.01)
22.5-27.5	(0.99-1.01)	(1.00-1.00)	(0.75-1.25)	(0.98-1.02)
27.5-32.5	(0.99-1.01)	(0.98-1.02)	(0.76-1.24)	(0.99-1.01)
32.5-37.5	(0.99-1.01)	(0.99-1.01)	(0.78-1.22)	(0.99-1.01)
37.5-42.5	(0.99-1.01)	(1.00-1.00)	(0.90-1.10)	(1.00-1.00)
42.5-47.5	(1.00-1.00)	(0.99-1.01)	(0.95-1.05)	(1.00-1.00)
52.5-57.5	(1.00-1.00)	(1.00-1.00)	(0.99-1.01)	(1.00-1.00)
57.5-62.5	(1.00-1.00)	(0.99-1.01)	(1.00-1.00)	(1.00-1.00)
62.5-67.5	(1.00-1.00)	(0.99-1.01)	(1.00-1.00)	(1.00-1.00)
67.5-72.5	(1.00-1.00)	(0.99-1.01)	(1.00-1.00)	(1.00-1.00)
72.5-77.5	(1.00-1.00)	(0.99-1.01)	(1.00-1.00)	(1.00-1.00)
77.5-82.5	(1.00-1.00)	(0.99-1.01)	(1.00-1.00)	(1.00-1.00)

Tabell 6.4.5: *Et 95% konfidensintervall på relativ endring i dødsintensiteten fra gjennomsnittlig dødsintensitet. Menn.*

6.4.3 Interpolering og bootstrapping

For mange av de sensurerte cellene har vi informasjon om antall døde i tidligere eller seinere år. Denne informasjonen kan benyttes til å foreta en form for kvalifisert gjetning på imputasjonsverdier.

De imputerte verdiene skal vi nå finne ved å interpolere mellom eventuelle ikke-imputerte verdier for død tidligere og seinere i perioden. La oss benevne de imputerte verdiene $\mathbf{X} = X_1, \dots, X_n$. For en vektor av dødsfall, $\{5, 0, X_i, 5, 7, 7, 8\}$, for en gitt uføreårsak, tilgangssår og kjønn, blir den ukjente verdien $X_i = (0 + 5)/2 = 2.5$. En slik metode gir for uføredatane høyere gjennomsnittlige imputeringsverdier enn 2.5. Disse listes i tabell 6.4.6.

Imputeringsverdiene $\mathbf{X} = X_1, \dots, X_n$ funnet ved interpolering antas som et forventnings-estimat, og benyttes videre til en parametrisk bootstrapping. For hver av de n imputerte cellene, trekkes m Poisson-fordelte variabler i henhold til disse forventningene. Variabler som ikke ligger mellom 1 og 4 forkastes, og nye variabler for disse trekkes til alle variablene er gyldige. Bootstrappingsforløpet skisseres som følger:

$$\underbrace{X_1, \dots, X_n}_{\mathbf{X}} \rightarrow \begin{matrix} X_{1,1}^* & , \dots, & X_{1,n}^* \\ \vdots & & \vdots \\ X_{m,1}^* & , \dots, & X_{m,n}^* \end{matrix} \rightarrow \underbrace{\bar{X}_1^*, \dots, \bar{X}_n^*}_{\mathbf{X}^*}$$

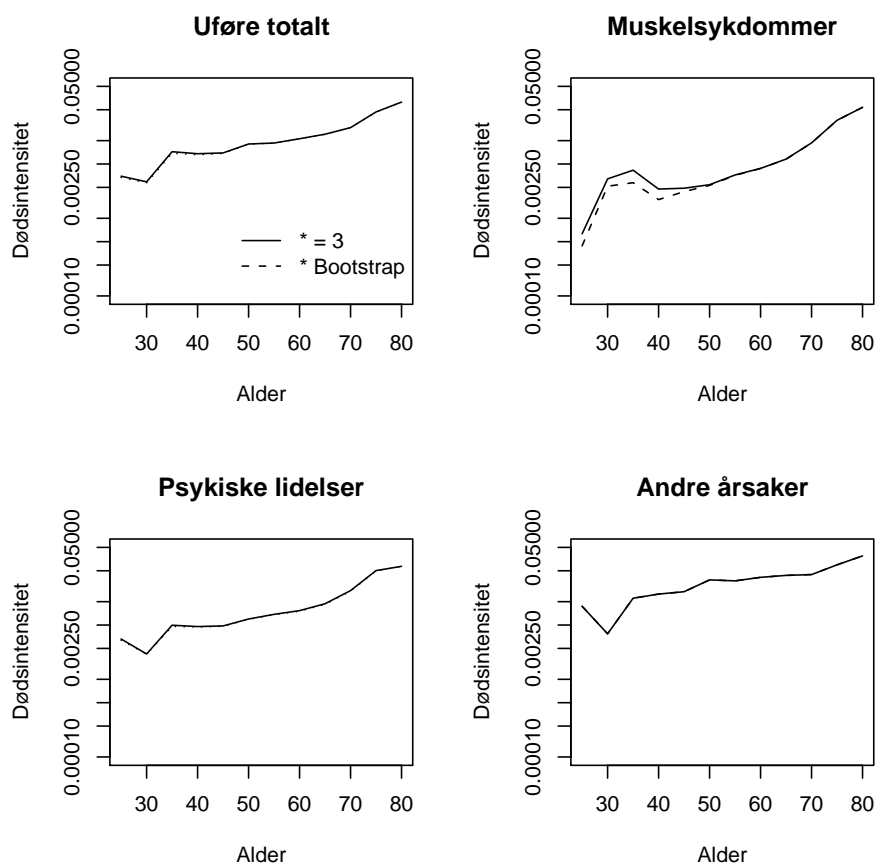
Forventningen til de interpolerte verdiene ($\bar{\mathbf{X}} = E[\mathbf{X}]$), samt forventning ($\bar{\mathbf{X}}^* = E[\mathbf{X}^*]$) og standardavvik ($\mathbf{X}_{sd}^* = sd[\mathbf{X}^*]$) til de bootstrappede verdiene er listet i tabell 6.4.6.

I figurene 6.4.5 og 6.4.6 vises dødsintensiteten beregnet ut ifra de bootstrappede verdiene og en fast verdi på 3 døde. Vi ser at dødsintensiteten for de muskelsyke blir lavere ved å benytte de bootstrappede verdiene enn en fast verdi på 3 døde. Forskjellene mellom de to imputeringsmetodene er, for de resterende uføreårsakene, så små at de ikke vises i figurene.

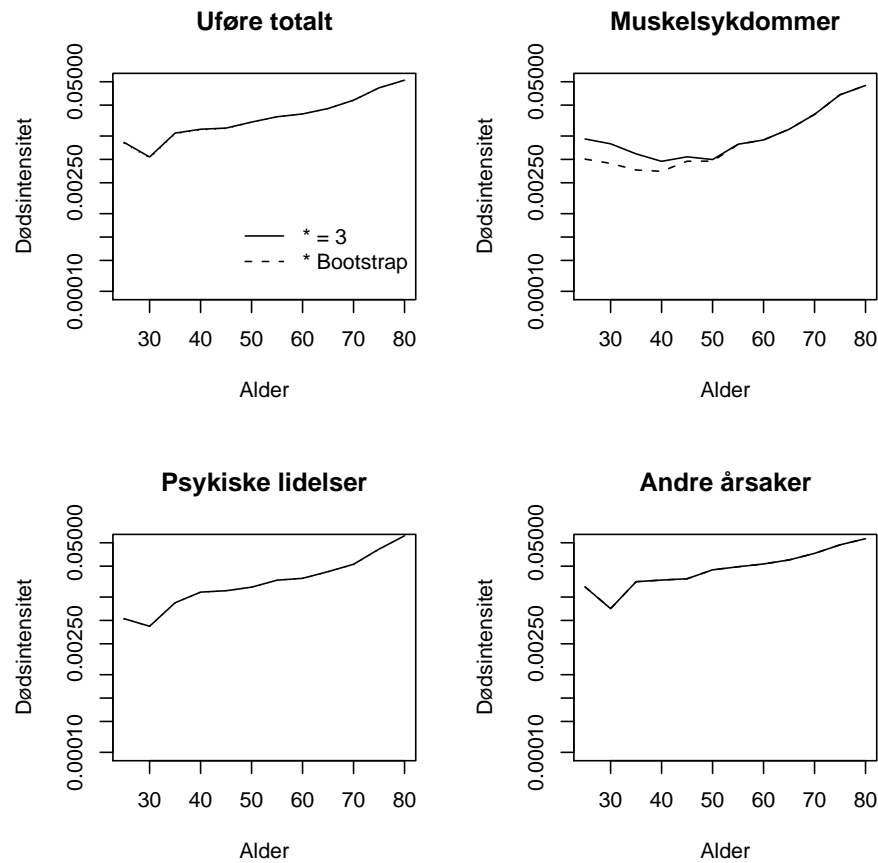
Tilnærmede 95% konfidensintervaller for endringene i dødsintensiteten gis i tabellene 6.4.7 og 6.4.8 for henholdsvis kvinner og menn. Variabiliteten i bootstrappingsresultatene viser seg å være liten for de uførepensjonerte grunnet psykiske lidelser og andre årsaker, men er i ung alder relativt stor for de uførepensjonerte grunnet muskelsykdommer.

Årsak	Kvinner		Menn	
	\bar{X}	$\bar{X}^*(X_{sd}^*)$	\bar{X}	$\bar{X}^*(X_{sd}^*)$
Psykiske lidelser	3.48	2.66(0.28)	3.69	2.74(0.28)
Muskelsykdommer	2.91	2.43(0.41)	2.66	2.33(0.41)
Andre årsaker	3.84	2.79(0.12)	4.00	2.84(0.12)

Tabell 6.4.6: *Imputasjonsverdier funnet ved interpolering og bootstrapping.*



Figur 6.4.5: *Dødsintensitet beregnet ut ifra interpolerte og bootstrappede imputasjonsverdier. Modell ADØD. Logaritmisk skala. Kvinner.*



Figur 6.4.6: Dødsintensitet beregnet ut ifra interpolerte og bootstrappede imputasjonsverdier. Modell ADØD. Logaritmisk skala. Menn.

Alder	Psykiske lidelser	Muskelsykdommer	Andre årsaker
17.5-22.5	(0.94 - 1.06)	(0.63 - 1.37)	(0.98 - 1.02)
22.5-27.5	(0.91 - 1.09)	(0.58 - 1.42)	(0.96 - 1.04)
27.5-32.5	(0.96 - 1.04)	(0.79 - 1.21)	(0.98 - 1.02)
32.5-37.5	(0.95 - 1.05)	(0.86 - 1.14)	(1.00 - 1.00)
37.5-42.5	(0.96 - 1.04)	(0.90 - 1.10)	(0.99 - 1.01)
42.5-47.5	(0.98 - 1.02)	(0.95 - 1.05)	(1.00 - 1.00)
52.5-57.5	(0.98 - 1.02)	(0.99 - 1.01)	(1.00 - 1.00)
57.5-62.5	(0.99 - 1.01)	(0.99 - 1.01)	(1.00 - 1.00)
62.5-67.5	(0.97 - 1.03)	(0.99 - 1.01)	(1.00 - 1.00)
67.5-72.5	(0.99 - 1.01)	(0.99 - 1.01)	(1.00 - 1.00)
72.5-77.5	(0.98 - 1.02)	(1.00 - 1.00)	(1.00 - 1.00)
77.5-82.5	(0.99 - 1.01)	(1.00 - 1.00)	(1.00 - 1.00)

Tabell 6.4.7: 95% konfidensintervall på relativ endring i dødsintensiteten fra gjennomsnittlig dødsintensitet for bootstrappede verdier. Kvinner

Alder	Psykiske lidelser	Muskelsykdommer	Andre årsaker
17.5-22.5	(0.91 - 1.09)	(0.60 - 1.40)	(0.98 - 1.02)
22.5-27.5	(0.93 - 1.07)	(0.63 - 1.37)	(0.95 - 1.05)
27.5-32.5	(0.94 - 1.06)	(0.81 - 1.19)	(0.98 - 1.02)
32.5-37.5	(0.96 - 1.04)	(0.78 - 1.22)	(0.99 - 1.01)
37.5-42.5	(0.97 - 1.03)	(0.90 - 1.10)	(0.99 - 1.01)
42.5-47.5	(0.98 - 1.02)	(0.96 - 1.04)	(1.00 - 1.00)
52.5-57.5	(0.98 - 1.02)	(0.98 - 1.02)	(1.00 - 1.00)
57.5-62.5	(0.97 - 1.03)	(0.99 - 1.01)	(1.00 - 1.00)
62.5-67.5	(0.97 - 1.03)	(0.99 - 1.01)	(1.00 - 1.00)
67.5-72.5	(0.98 - 1.02)	(0.99 - 1.01)	(1.00 - 1.00)
72.5-77.5	(0.97 - 1.03)	(1.00 - 1.00)	(1.00 - 1.00)
77.5-82.5	(0.99 - 1.01)	(1.00 - 1.00)	(1.00 - 1.00)

Tabell 6.4.8: 95% konfidensintervall på relativ endring i dødsintensiteten fra gjennomsnittlig dødsintensitet for bootstrappede verdier. Menn.

6.5 Vurdering av sensitivitet

Det er tydelig at det finnes mange kilder til feilestimering av dødsintensitetene i denne oppgaven. Det virker imidlertid som at de fleste feilkildene tilfører usikkerhet av en begrenset størrelse. Ved ikke å ta hensyn til immigrasjon og emigrasjon, ser vi at dødsintensiteten for befolkningen endres noe, men av svært liten betydning. Det samme gjelder aggregeringen av dataene. Her får vi noe avvikende resultater i ung alder, men ellers ser de to metodene ut til å gi ensbetydende resultater.

Antagelsene om en Gompertz-fordeling gir store feilestimeringer i ung alder, siden dødsintensiteten ikke er log-lineær gjennom hele aldersperspektivet. Vi ser også at regresjonen gir en bedre, glattet tilpasning til de årlige ratene, enn hva Gompertz-fordelingen gjør. Spesielt viktig er det at den fanger opp avvikene fra log-linearitet i ung alder.

Den siste feilkilden som har blitt vurdert, er valg av imputasjonsverdier for de sensurerte cellene. Fra figurene som viser maksimal spredning på imputasjonsverdiene, ser vi at den største usikkerheten forekommer i ung alder, hvor vi observerer få dødsfall. Usikkerheten som foreligger på grunn av valg av imputasjonsverdier er størst for gruppa av uførepensjonerte grunnet muskelsykdommer. Imidlertid ser det ikke ut til at usikkerheten tilknyttet imputasjonsverdiene skaper noe videre problemer i sammenligningen med befolkningsdødeligheten. Det meste av imputeringen foretas i ung alder, hvor vi allerede har en betydelig overdødelighet blant samtlige uføregrupper.

Kapittel 7

Diskusjon

7.1 Oppsummering av resultater

På grunn av formen dataene foreligger på, var store endringer i dataformatet påkrevd. Det var samtidig nødvendig å foreta antagelser om hvilken alder individene hadde ved tilgang, samt tiden de var til risiko i studien.

I kapittel 4 tilpasset vi en Gompertz-fordeling til befolkningsdataene. Tilpasningen var utilstrekkelig i ung alder hvor dødsintensitetskurven ikke hadde en log-lineær form. Dette kan imidlertid tilpasses bedre ved å benytte en Gompertz-Makeham-fordeling, der et konstantledd vil fange opp denne effekten. Selv om tilpasningen av Gompertz-fordelingen ikke er optimal, er fordelingen lett å anvende. Kun to parametere tilpasses. Disse er relativt intuitive, og gir oss god kontroll over modellen. Med en kjent fordelingsfunksjon kan vi også predikere dødsintensiteten framover i tid.

Vi har funnet store forskjeller i dødsintensitetene mellom de ulike årsakene til uførhet, og i sammenligning med den generelle befolkningsdødeligheten. Det viser seg at disse forskjellene er sterkt aldersbetinget. Dødsintensiteten er signifikant høyere blant de uføre enn for befolkningen. For de uføretrygdde grunnet psykiske lidelser kunne vi ikke påvise signifikante forskjeller i dødsintensiteten mot befolkningen i høye aldre. Det viser seg også at dødsintensiteten i høye aldre blant de uførepensjonerte grunnet muskelsykdommer er signifikant lavere enn for den generelle befolkningen.

I kapittel 5 benyttet vi Poisson-regresjon til tilpasningen av dødsintensitetene. Dette gav oss generelt bedre tilpasninger. Vi så hvor mange og hvilke forklaringsvariabler som var nødvendige for å få en god tilpasning av dødsintensiteten for de ulike uføreårsakene. Effekten av de ulike forklaringsvariablene ble også vurdert. Vi fant at dødsintensiteten best beskrives av forklaringsvariabelen alder ved død, sammen med med tid fra tilgang og tilgangsåret. Fra forklaringsvariabelen tilgangsårså så vi en generell reduksjon i dødsintensiteten etter som årene for tilgang til uførepensjon økte. For de uførepensjonerte grunnet andre årsaker så vi en kraftig økning i dødsintensiteten for tilgang i 2004-2005. Vi kan muligens relatere dette til ordningen med tidsbegrenset uførestønad som ble innført 1.1.2004. De med tidsbegrenset uførestønad er ikke inkludert i dataene.

Da vi så på effekten av forklaringsvariabelen tid fra tilgang, fant vi også bemerkelses-

verdige forskjeller mellom uføreårsakene. For de uføre grunnet psykiske lidelser og muskelsykdommer, er dødsintensiteten jevnt stigende. Gruppen av uførepensjonerte grunnet andre årsaker har høyest dødsintensitet de første årene etter tilgang til uførhet. Deretter stabiliserer dødsintensiteten seg på et lavere nivå etter noen år.

Det at dataene foreligger på aggregert form, og at det eksakte antallet døde i noen tilfeller er ukjent, gjør at beregningene som foretas blir usikre. En rekke sensitivitetsvurderinger har blitt foretatt i kapittel 6. Ut i fra dette har vi fått et visst begrep om størrelsen på usikkerheten i de enkelte tilfellene. Allikevel vil den totale usikkerheten som foreligger på grunn av aggregeringen av dataene, svekke resultatene.

7.2 Tidligere studier

Som nevnt innledningsvis er det et begrenset antall studier som har blitt foretatt om dødeligheten blant uførepensjonerte. Spesielt lite årsaksrelaterte studier er funnet. I det følgende nevnes et knippe av sammenlignbare studier.

Paulsboe viser i sin masteroppgave[10] også økt dødelighet blant de uførepensjonerte. Spesielt påviser han en lavere dødelighet blant gruppa uførepensjonert grunnet muskelskjelettsykdommer enn for befolkningen. Dette gjelder for kvinner fra 60 år og menn fra 65. Se Paulsboe[10] kap. 4.10. Eksakte tall på dette, eller hvorvidt den lavere dødeligheten er signifikant eller ikke, framgår ikke av oppgaven. Studien er basert på et datamateriale over alle uføre i Norge i alderen 20-67 i perioden fra 1970 til 1990. Datamaterialet som er benyttet i Paulsboes oppgave er altså ikke overlappende med dataene fra denne oppgaven.

I Karlsson et al.[7] baseres studien på en 12-årig oppfølging av personer fra det svenske fylket Östergötland i perioden fra 1985 til 1996. Studien viser signifikant høyere dødelighet blant de uførepensjonerte enn for befolkningen. Overdødeligheten viser seg også å være synkende med alder. Studien skiller imidlertid ikke uførepensjonistene inn i ulike årsaker til uførhet.

En tilsvarende studie er foretatt for langtidssykemeldte i Gjesdal et al. [6] hvor undersøkelsen baseres på befolkningen i Hordaland i årene fra 1994 til 2003. Også for de sykemeldte viser dødeligheten seg å være signifikant høyere enn for befolkningen. Gruppen av sykemeldte grunnet muskelsykdommer skilles ut, og disse gir blant kvinnene et standardisert dødelighetsforhold mindre enn 1. Forholdet er imidlertid ikke signifikant lavere.

Under arbeidet med denne oppgaven ble nok en artikkel om emnet publisert av Gjesdal et al. [5]. Denne artikkelen baserer seg på data fra et 10% randomisert utvalg av Norges befolkning i alderen 30-59 år og for alle innbyggere i alderen 30-66 år i Östergötland. Begge populasjonene ble fulgt opp i perioden fra 1990 til 1996. I denne artikkelen finner vi de samme overdødelighetene blant de uføre som vi har sett tidligere. Her er også gruppa av uføre splittet opp i ulike årsaker til uførhet. Gruppen som ble uførepensjonert grunnet muskelsykdommer viser seg å ikke ha signifikant høyere dødelighet enn befolkningen. Dette gjelder for begge populasjonene. Gruppen av psykiske lidelser er i artikkelen av Gjesdal et al.[5] splittet opp i ulike undergrupper av årsaker. Disse har en signifikant høyere dødelighet enn befolkningen. Ytterligere beregninger må til for å kunne sammenligne resultatene fra artikkelen til Gjesdal et al.[5] mot resultatene i denne oppgaven. Man må være oppmerksom på at

datamaterialene spanner over ulike perioder og aldre for de uføre.

I artikkelen av Gjesdal et al.[5] estimeres konstante hazardratioer for død i oppfølgings-tiden. Som vi har sett igjennom denne oppgaven, er den relative risikoen for død sterkt aldersbetinget. Dette betyr at stratifiserte analyser for ulike aldersgrupper bør foretas.

7.3 Konklusjon

Vi har i oppgaven observert endel interessante effekter i dødsintensiteten til de uføre. Først og fremst så vi signifikant høyere dødsintensitet blant de uføre enn blant resten av befolkningen. Dette tilsvarte hva resultatene fra andre studier også viste. Vi fant signifikant lavere dødsintensiteter blant gruppa uførepensjonert grunnet muskelsykdommer i høy alder. Tilsvarende, men ikke signifikante, tendenser har vi også sett i tidligere studier. Studiene vi har sammenlignet med baseres imidlertid på andre datagrunnlag og andre modeller er også benyttet til tilpasningen av dataene. På grunn av den aggregerte formen dataene foreligger på, har analysene blitt tilknyttet en del usikkerhetsmomenter, som svekker styrken i resultatene. Dette gjør at det er ønskelig å foreta de samme analysene med data på individnivå for å verifisere resultatene.

Tillegg A

Beregning av forventede observasjonstider

For å beregne de forventede gjennomlevde personårene fra tilgang til uførepensjon og til død, benyttes en liten Monte-Carlo-simulering. Et tilfeldig tidspunkt trekkes uniformt i tilgangsårsgruppa og deretter trekkes et tidspunkt fra dødsårsgruppa. Avstanden mellom disse tidene kaller vi observasjonstider gitt død. Gitt overlevelse gjennom hele intervallet beregnes avstanden fram til dødsårsgruppas øvre grense.

Det forekommer også tilfeller hvor tilgangsårs- og dødsårsgruppa delvis overlapper hverandre. Da trekkes dødstidspunktet fra et intervall mellom maksimum av trukket tilgangsårs og nedre grense for dødsårsgruppa, og til øvre grense for dødsårsgruppa. Dette gjentas en rekke ganger før et gjennomsnitt av observasjonstider gitt død og observasjonstider gitt overlevelse blir funnet.

For personer som har tilgang i årene 1992-1993 og dør i årene 1992-1994, blir den gjennomsnittlige levetiden 1 år, slik R-koden viser:

```
aar      <- 1992
dod_start <- 1992
dod_stopp <- 1995

sim  <- 100000
tilg <- runif(sim,aar,(aar+2))
dod  <- runif(sim,pmax(tilg,dod_start),dod_stopp)

mean(dod_stopp-pmax(tilg,dod_start))
mean(dod-pmax(tilg,dod_start))

> mean(dod_stopp-pmax(tilg,dod_start))
[1] 1.998301

> mean(dod-pmax(tilg,dod_start))
[1] 1.001210
```

Tillegg A. Beregning av forventede observasjonstider

For aggregerte befolkningsdata blir dette imidlertid noe annerledes. Siden vi vet befolkningsstørrelsen per 1.1.1992 og 1.1.1993 trekkes tilgangsåret binomisk, og en forventet levetid beregnes ut ifra denne.

```
aar      <- 1992
dod_start <- 1992
dod_stopp <- 1995

sim      <- 100000
tilg     <- aar + rbinom(sim,1,0.5)
dod      <- runif(sim,pmax(tilg,dod_start),dod_stopp)

mean(dod_stopp-pmax(tilg,dod_start))
mean(dod-pmax(tilg,dod_start))

> mean(dod_stopp-pmax(tilg,dod_start))
[1] 2.50104
> mean(dod-pmax(tilg,dod_start))
[1] 1.252097
```

Tillegg B

Konfidensintervaller

For å utlede konfidensintervaller for $\hat{\mu}_j$ og $t\hat{p}_x$ trenger vi et teorem om konvergens i fordeling. Se Lehmann[8] kap. 2.5.

Teorem B.0.1 (δ -metoden) *Hvis*

$$\sqrt{n}(T_n - \xi) \xrightarrow{\mathcal{D}} N(0, \tau^2)$$

så vil

$$\sqrt{n}(g(T_n) - g(\xi)) \xrightarrow{\mathcal{D}} N(0, \tau^2[g'(\xi)]^2)$$

dersom $g(\xi)$ eksisterer og er forskjellig fra null.

B.1 Konfidensintervaller for $\hat{\mu}$

La oss anta at følgen T_n er vår følge av stykkevis konstante dødsintensiteter $\hat{\mu}_j$, og at $g(\hat{\mu}_j) = \log(\hat{\mu}_j)$. Vi kan dermed finne estimat på forventningen og variansen til $\log(\hat{\mu}_j)$. Det kan vises at $\log(\hat{\mu}_j)$ er tilnærmet forventningsrett og at variansen blir

$$\text{Var}[\log(\hat{\mu}_j)] \approx \left[\frac{\partial}{\partial \mu_j} f(\hat{\mu}_j) \right]^2 \text{Var}[\hat{\mu}_j].$$

Et estimat for variansen til $\log(\hat{\mu}_j)$ blir dermed

$$\frac{D_j}{T_j^2} \frac{T_j^2}{D_j^2} = \frac{1}{D_j}.$$

Ved transformasjon vil et tilnærmet $(1 - \alpha)100\%$ konfidensintervall for $\hat{\mu}_j$ være

$$\frac{D_j}{T_j} \exp \left(\pm z(\alpha/2) \sqrt{\frac{1}{D_j}} \right), \quad (\text{B.1})$$

hvor $z(\alpha/2)$ er α -kvantilen i en standard normalfordeling.

B.2 Konfidensintervaller for ${}_t\hat{p}_x$

Fra (3.3) er overlevelsessannsynligheten definert ved kontinuerlig tid. Vi antar nå stykkevis konstante dødsintensiteter, jamfør definisjon 3.1.2. Overlevelsessannsynligheten vet vi fra (3.9) at da kan skrives som

$$\begin{aligned} {}_t\hat{p}_x &= \exp\left(-\sum_{k=0}^t \hat{\mu}_{j-k}\right) \\ &= \exp\left(-\sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}}\right). \end{aligned}$$

${}_t\hat{p}_x$ kan vises å være tilnærmet forventningsrett. Fra kovariansmatrisen $\bar{\mathbf{J}}^{-1}(\hat{\boldsymbol{\mu}})$, kjenner vi variansen til $\hat{\mu}_j$. Dermed kan variansen til ${}_t\hat{p}_x$ utledes. Utledningen tilsvarer den for $\hat{\mu}_j$ ved først å finne variansen til $\log {}_t\hat{p}_x$ og deretter benytte teorem B.0.1, med $g(\log {}_t\hat{p}_x) = \exp(\log {}_t\hat{p}_x)$. Vi har da at

$$\begin{aligned} \text{Var}[\log {}_t\hat{p}_x] &= \text{Var}\left[\log\left(\exp\left(-\sum_{k=0}^t \hat{\mu}_{j-k}\right)\right)\right] \\ &= \sum_{k=0}^t \text{Var}[\hat{\mu}_{j-k}], \end{aligned}$$

hvor siste likhet følger siden $\{\bar{\mathbf{J}}^{-1}(\hat{\boldsymbol{\mu}})\}_{ij} = 0$, når $i \neq j$. Dermed har vi at

$$\begin{aligned} \text{Var}[_t\hat{p}_x] &\approx \left[\frac{\partial}{\partial \log {}_t\hat{p}_x} \exp(\log {}_t\hat{p}_x)\right]^2 \sum_{k=0}^t \text{Var}[\hat{\mu}_{j-k}] \\ &= [({}_t\hat{p}_x)]^2 \sum_{k=0}^t \text{Var}[\hat{\mu}_{j-k}]. \end{aligned}$$

Innsatt de årlige ratene, vil et estimat på variansen være

$$\exp\left(-\sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}}\right)^2 \sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}^2}.$$

Et tilnærmet $(1 - \alpha)100\%$ konfidensintervall for ${}_t\hat{p}_x$ vil med dette være

$$\exp\left(-\sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}}\right) \pm z(\alpha/2) \exp\left(-\sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}}\right) \sqrt{\sum_{k=0}^t \frac{D_{j-k}}{T_{j-k}^2}}. \quad (\text{B.2})$$

B.3 Konfidensintervaller for \tilde{c}

Vi benytter tilsvarende metode for utledningen av dette konfidensintervallet som for konfidensintervallet til $\hat{\mu}_j$. $\log(\tilde{c}_j)$ er tilnærmet forventningsrett. For variansen har vi at

$$\text{Var} \left(\log \left(\frac{\hat{\mu}_{j+1}}{\hat{\mu}_j} \right) \right) = \text{Var} (\log(\hat{\mu}_{j+1})) + \text{Var} (\log(\hat{\mu}_j)).$$

Estimatet på variansen til $\log(\hat{\mu}_j)$ ble utledet i tillegg B.1, og ved å benytte dette får vi at et estimat på variansen til \tilde{c} vil være

$$\frac{1}{D_{j+1}} + \frac{1}{D_j}.$$

Ved transformasjon vil dermed et tilnærmet $(1 - \alpha)100\%$ konfidensintervall for \tilde{c} være gitt ved

$$\left(\frac{T_j D_{j+1}}{D_j T_{j+1}} \right) \exp \left(\pm z(\alpha/2) \left(\sqrt{\frac{1}{D_{j+1}} + \frac{1}{D_j}} \right) \right).$$

Bibliografi

- [1] O. A. AALEN, ØRNULF BORGAN, AND H. K. GJESSING, *Survival and Event History Analysis*, Springer, 2008.
- [2] ARBEIDS- OG INKLUDERINGSDEPARTEMENTET, *Lov om folketrygd*, 1997.
- [3] L. FAHRMEIR AND G. TUTZ, *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer, second ed., 2001.
- [4] H. GERBER, *Life Insurance Mathematics*, Springer, third ed., 1997.
- [5] S. GJESDAL, P. R. RINGDAL, K. HAUG, J. G. MAELAND, S. E. VOLLSET, AND K. A. E. ALEXANDERSON, *Mortality after long-term sickness absence: prospective cohort study*, Euroean Journal of Public Health, 18 (2008), pp. 517–521.
- [6] S. GJESDAL, P. SVEDBERG, J. HAGBERG, AND K. A. E. ALEXANDERSON, *Mortality in relation to disability pensioners in norway and sweeden 1990-96: Comparative prospective cohort study*, Scandinavian Journal of Public Health, 37 (2009), pp. 168–175.
- [7] N. E. KARLSON, J. M. CARSTENSEN, S. GJESDAL, AND K. A. E. ALEXANDERSON, *Mortality in relation to disability pension: Findings from a 12-year prospective population-bases cohort study in sweden*, Scandinavian Journal of Public Health, 35 (2007), pp. 341–347.
- [8] E. L. LEHMANN, *Elements of Large-Sample Theory*, Springer, 1999.
- [9] P. McCULLAGH AND J. NELDER, *Generalized Linear Models*, Chapman and Hall, second ed., 1989.
- [10] A. E. PAULSBOE, *En statistisk analyse av dødeligheten blant uføretrygdde i norge 1970-1990*, 1997.
- [11] R DEVELOPMENT CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2007.